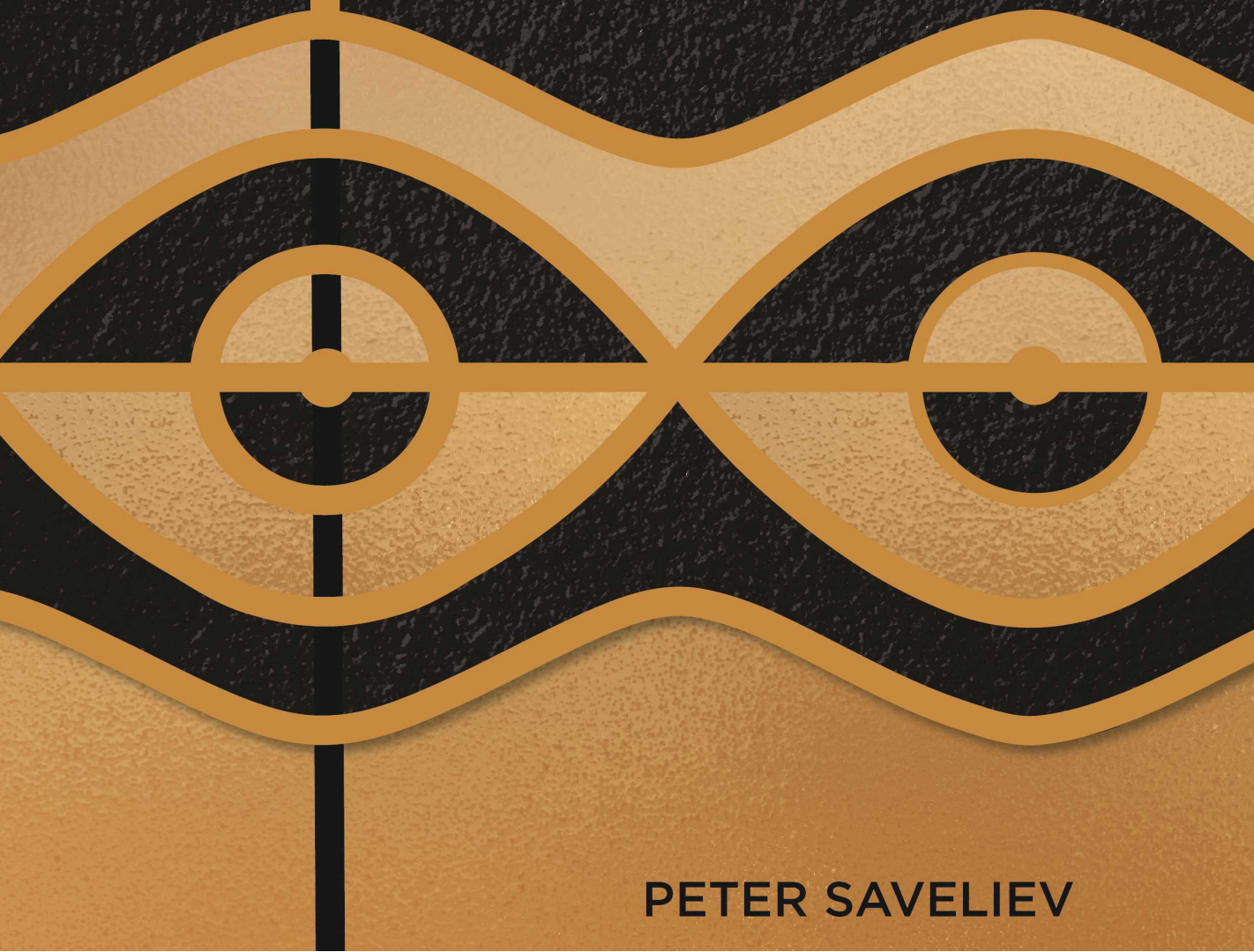


CALCULUS ILLUSTRATED

VOLUME 5:
DIFFERENTIAL
EQUATIONS



PETER SAVELIEV

To the student

Mathematics is a science. Just as the rest of the scientists, mathematicians are trying to understand how the Universe operates and discover its laws. When successful, they write these laws as short statements called “theorems”. In order to present these laws conclusively and precisely, a dictionary of the new concepts is also developed; its entries are called “definitions”. These two make up the most important part of any mathematics book.

This is how definitions, theorems, and some other items are used as building blocks of the scientific theory we present in this text.

Every new concept is introduced with utmost specificity.

Definition 0.0.1: square root

Suppose a is a positive number. Then the *square root* of a is a positive number x , such that $x^2 = a$.

The term being introduced is given in *italics*. The definitions are then constantly referred to throughout the text.

New symbolism may also be introduced.

Square root

\sqrt{a}

Consequently, the notation is freely used throughout the text.

We may consider a specific instance of a new concept either before or after it is explicitly defined.

Example 0.0.2: length of diagonal

What is the length of the diagonal of a 1×1 square? The square is made of two right triangles and the diagonal is their shared hypotenuse. Let’s call it a . Then, by the *Pythagorean Theorem*, the square of a is $1^2 + 1^2 = 2$. Consequently, we have:
$$a^2 = 2.$$
We immediately see the need for the square root! The length is, therefore, $a = \sqrt{2}$.

You can skip some of the examples without violating the flow of ideas, at your own risk.

All new material is followed by a few little tasks, or questions, like this.

Exercise 0.0.3

Find the height of an equilateral triangle the length of the side of which is 1.

The exercises are to be attempted (or at least considered) immediately.

Most of the in-text exercises are not elaborate. They aren’t, however, entirely routine as they require understanding of, at least, the concepts that have just been introduced. Additional exercise *sets* are placed in the appendix as well as at the book’s website: calculus123.com. Do not start your study with the exercises! Keep in mind that the exercises are meant to test – indirectly and imperfectly – how well the *concepts* have been learned.

There are sometimes words of caution about common mistakes made by the students.

Warning!

In spite of the fact that $(-1)^2 = 1$, there is only one square root of 1, $\sqrt{1} = 1$.

The most important facts about the new concepts are put forward in the following manner.

Theorem 0.0.4: Product of Roots

For any two positive numbers a and b , we have the following identity:

$$\sqrt{a} \cdot \sqrt{b} = \sqrt{a \cdot b}$$

The theorems are constantly referred to throughout the text.

As you can see, theorems may contain formulas; a theorem supplies limitations on the applicability of the formula it contains. Furthermore, every formula is a part of a theorem, and using the former without knowing the latter is perilous.

There is no need to memorize definitions or theorems (and formulas), initially. With enough time spent with the material, the main ones will eventually become familiar as they continue to reappear in the text. Watch for words “important”, “crucial”, etc. Those new concepts that do not reappear in this text are likely to be seen in the next mathematics book that you read. You need to, however, be aware of all of the definitions and theorems and be able to find the right one when necessary.

Often, but not always, a theorem is followed by a thorough argument as a justification.

Proof.

Suppose $A = \sqrt{a}$ and $B = \sqrt{b}$. Then, according to the [definition](#), we have the following:

$$a = A^2 \text{ and } b = B^2 .$$

Therefore, we have:

$$a \cdot b = A^2 \cdot B^2 = A \cdot A \cdot B \cdot B = (A \cdot B) \cdot (A \cdot B) = (AB)^2 .$$

Hence, $\sqrt{ab} = A \cdot B$, again according to the definition.

Some proofs can be skipped at first reading.

Its highly detailed exposition makes the book a good choice for *self-study*. If this is your case, these are my suggestions.

While reading the book, try to make sure that you understand new concepts and ideas. Keep in mind, however, that some are more important than others; they are marked accordingly. Come back (or jump forward) as needed. Contemplate. Find other sources if necessary. You should not turn to the exercise sets until you have become comfortable with the material.

What to do about exercises when solutions aren’t provided? First, use the examples. Many of them contain a problem – with a solution. Try to solve the problem – before or after reading the solution. You can also find exercises online or make up your own problems and solve them!

I strongly suggest that your solution should be thoroughly *written*. You should write in complete sentences, including all the algebra. For example, you should appreciate the difference between these two:

Wrong: $\frac{1+1}{2}$

Right: $\frac{1+1}{=2}$

The latter reads “one added to one is two”, while the former cannot be read. You should also justify all your steps and conclusions, including all the algebra. For example, you should appreciate the difference between these two:

Wrong:

$$\begin{array}{l} 2x = 4 \\ x = 2 \end{array}$$

Right:

$$\begin{array}{l} 2x = 4; \text{ therefore,} \\ x = 2. \end{array}$$

The standards of thoroughness are provided by the examples in the book.

Next, your solution should be thoroughly *read*. This is the time for self-criticism: Look for errors and weak spots. It should be re-read and then rewritten. Once you are convinced that the solution is correct and the presentation is solid, you may show it to a knowledgeable person for a once-over.

Next, you may turn to modeling projects. Spreadsheets (Microsoft Excel or similar) are chosen to be used for graphing and modeling. One can achieve as good results with packages specifically designed for these purposes, but spreadsheets provide a tool with a wider scope of applications. Programming is another option.

Good luck!

August 10, 2020

To the teacher

The bulk of the material in the book comes from my lecture notes.

There is little emphasis on closed-form computations and algebraic manipulations. I do think that a person who has never integrated by hand (or differentiated, or applied the quadratic formula, etc.) cannot possibly understand integration (or differentiation, or quadratic functions, etc.). However, a large proportion of time and effort can and should be directed toward:

- understanding of the concepts and
- modeling in realistic settings.

The challenge of this approach is that it requires more abstraction rather than less.

Visualization is the main tool used to deal with this challenge. Illustrations are provided for every concept, big or small. The pictures that come out are sometimes very precise but sometimes serve as mere metaphors for the concepts they illustrate. The hope is that they will serve as visual “anchors” in addition to the words and formulas.

It is unlikely that a person who has never plotted the graph of a function by hand can understand graphs or functions. However, what if we want to plot more than just a few points in order to visualize curves, surfaces, vector fields, etc.? Spreadsheets were chosen over graphic calculators for visualization purposes because they represent the shortest step away from pen and paper. Indeed, the data is plotted in the simplest manner possible: one cell - one number - one point on the graph. For more advanced tasks such as modeling, spreadsheets were chosen over other software and programming options for their wide availability and, above all, their simplicity. Nine out of ten, the spreadsheet shown was initially created from scratch in front of the students who were later able to follow my footsteps and create their own.

About the tests. The book isn't designed to prepare the student for some preexisting exam; on the contrary, assignments should be based on what has been learned. The students' understanding of the concepts needs to be tested but, most of the time, this can be done only indirectly. Therefore, a certain share of routine, mechanical problems is inevitable. Nonetheless, no topic deserves more attention just because it's likely to be on the test.

If at all possible, don't make the students memorize formulas.

In the order of topics, the main difference from a typical calculus textbook is that sequences come before everything else. The reasons are the following:

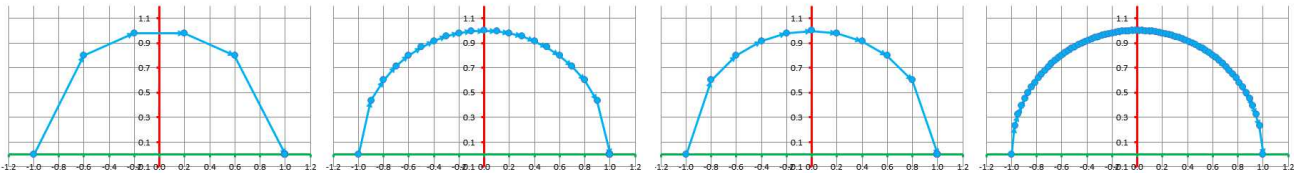
- Sequences are the simplest kind of functions.
- Limits of sequences are simpler than limits of general functions (including the ones at infinity).
- The sigma notation, the Riemann sums, and the Riemann integral make more sense to a student with a solid background in sequences.
- A quick transition from sequences to series often leads to confusion between the two.
- Sequences are needed for modeling, which should start as early as possible.

From the discrete to the continuous

It’s no secret that a vast majority of calculus students will never use what they have learned. Poor career choices aside, a former calculus student is often unable to recognize the mathematics that is supposed to surround him. Why does this happen?

Calculus is the science of change. From the very beginning, its peculiar challenge has been to study and measure *continuous* change: curves and motion along curves. These curves and this motion are represented by *formulas*. Skillful manipulation of those formulas is what solves calculus problems. For over 300 years, this approach has been extremely successful in sciences and engineering. The successes are well-known: projectile motion, planetary motion, flow of liquids, heat transfer, wave propagation, etc. Teaching calculus follows this approach: An overwhelming majority of what the student does is manipulation of formulas on a piece of paper. But this means that all the problems the student faces were (or could have been) solved in the 18th or 19th centuries!

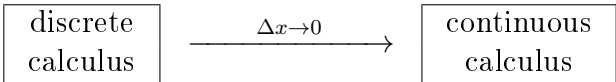
This isn’t good enough anymore. What has changed since then? The computers have appeared, of course, and computers don’t manipulate formulas. They don’t help with solving – in the traditional sense of the word – those problems from the past centuries. Instead of *continuous*, computers excel at handling *incremental* processes, and instead of formulas they are great at managing discrete (digital) data. To utilize these advantages, scientists “discretize” the results of calculus and create algorithms that manipulate the digital data. The solutions are approximate but the applicability is unlimited. Since the 20th century, this approach has been extremely successful in sciences and engineering: aerodynamics (airplane and car design), sound and image processing, space exploration, structure of the atom and the universe, etc. The approach is also circuitous: Every concept in calculus *starts* – often implicitly – as a discrete approximation of a continuous phenomenon!



Calculus is the science of change, *both* incremental and continuous. The former part – the so-called discrete calculus – may be seen as the study of incremental phenomena and the quantities *indivisible* by their very nature: people, animals, and other organisms, moments of time, locations of space, particles, some commodities, digital images and other man-made data, etc. With the help of the calculus machinery called “limits”, we invariably choose to transition to the continuous part of calculus, especially when we face continuous phenomena and the quantities *infinitely divisible* either by their nature or by assumption: time, space, mass, temperature, money, some commodities, etc. Calculus produces definitive results and absolute accuracy – but only for problems amenable to its methods! In the classroom, the problems are simplified until they become manageable; otherwise, we circle back to the discrete methods in search of approximations.

Within a typical calculus course, the student simply never gets to complete the “circle”! Later on, the graduate is likely to think of calculus only when he sees formulas and rarely when he sees numerical data.

In this book, every concept of calculus is first introduced in its discrete, “pre-limit”, incarnation – elsewhere typically hidden inside proofs – and then used for modeling and applications well before its continuous counterpart emerges. The properties of the former are discovered first and then the matching properties of the latter are found by making the increment smaller and smaller, at the *limit*:



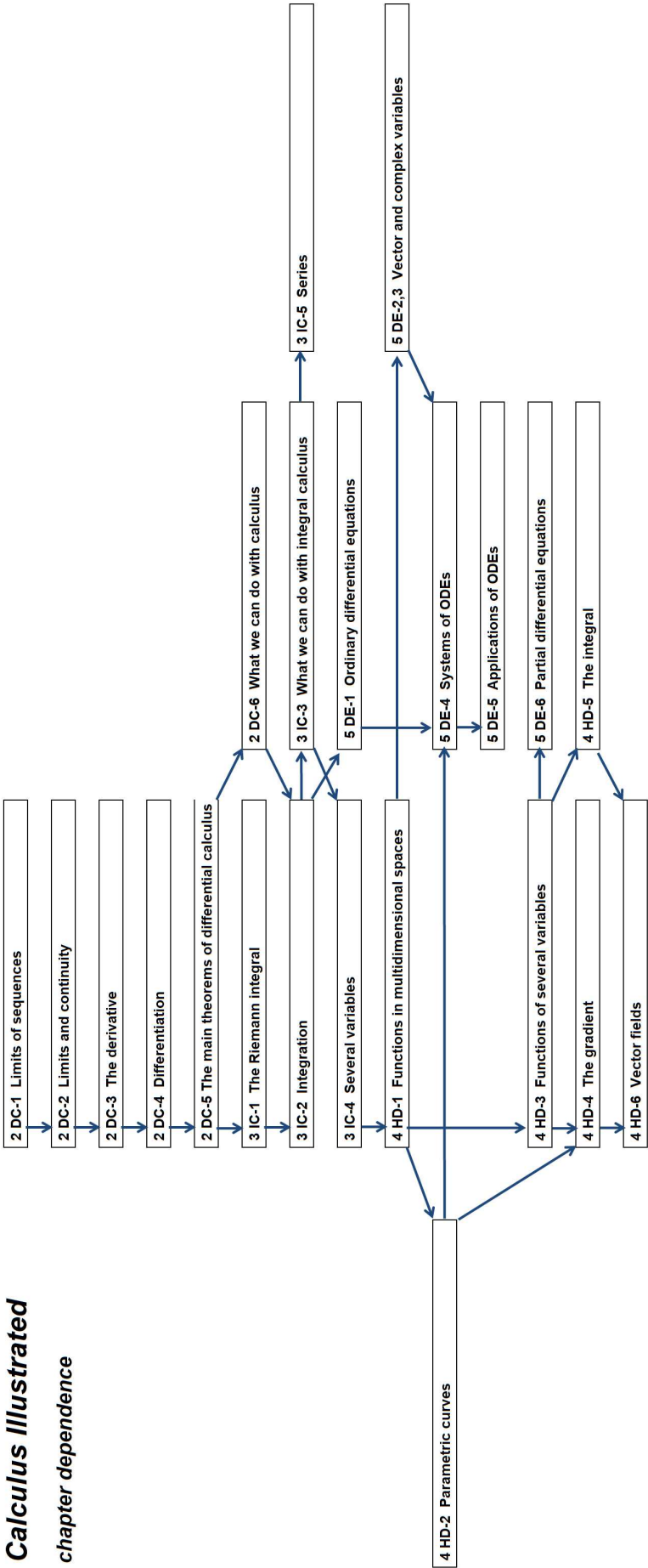
The volume and chapter references for *Calculus Illustrated*

This book is a part of the series *Calculus Illustrated*. The series covers the standard material of the undergraduate calculus with a substantial review of precalculus and a preview of elementary ordinary and partial differential equations. Below is the list of the books of the series, their chapters, and the way the present book (parenthetically) references them.

1 PC-1 1 PC-2 1 PC-3 1 PC-4 1 PC-5	■ Calculus Illustrated. Volume 1: Precalculus Calculus of sequences Sets and functions Compositions of functions Classes of functions Algebra and geometry
2 DC-1 2 DC-2 2 DC-3 2 DC-4 2 DC-5 2 DC-6	■ Calculus Illustrated. Volume 2: Differential Calculus Limits of sequences Limits and continuity The derivative Differentiation The main theorems of differential calculus What we can do with calculus
3 IC-1 3 IC-2 3 IC-3 3 IC-4 3 IC-5	■ Calculus Illustrated. Volume 3: Integral Calculus The Riemann integral Integration What we can so with integral calculus Several variables Series
4 HD-1 4 HD-2 4 HD-3 4 HD-4 4 HD-5 4 HD-6	■ Calculus Illustrated. Volume 4: Calculus in Higher Dimensions Functions in multidimensional spaces Parametric curves Functions of several variables The gradient The integral Vector fields
5 DE-1 5 DE-2 5 DE-3 5 DE-4 5 DE-5 5 DE-6	■ Calculus Illustrated. Volume 5: Differential Equations Ordinary differential equations Vector variables Vector and complex variables Systems of ODEs Applications of ODEs Partial differential equations

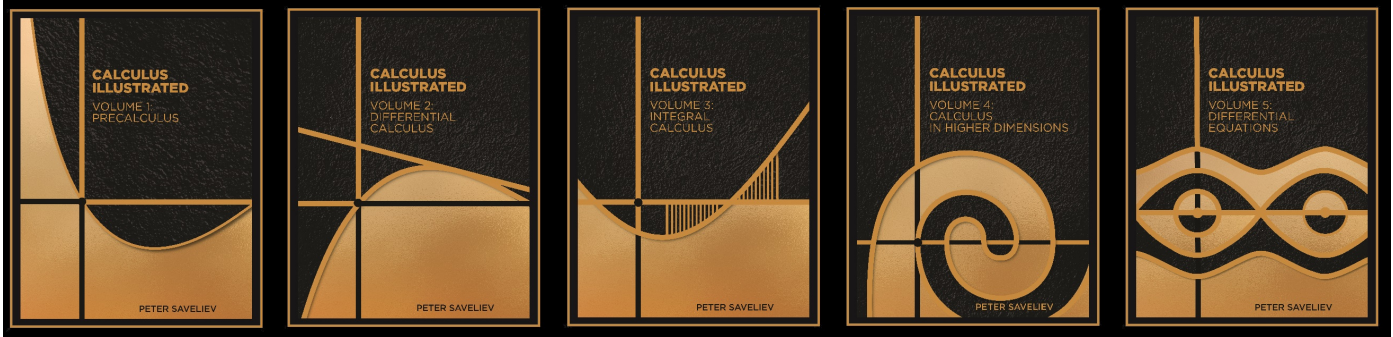
Each volume can be read independently.

A possible sequence of chapters is presented below. An arrow from A to B means that chapter B shouldn't be read before chapter A.



About the author

Peter Saveliev is a professor of mathematics at Marshall University, Huntington, West Virginia, USA. After a Ph.D. from the University of Illinois at Urbana-Champaign, he devoted the next 20 years to teaching mathematics. Peter is the author of a graduate textbook *Topology Illustrated* published in 2016. He has also been involved in research in algebraic topology and several other fields. His non-academic projects have been: digital image analysis, automated fingerprint identification, and image matching for missile navigation/guidance.



Contents

Preface	2
■ Chapter 1: Ordinary differential equations	12
1.1 Incremental motion	12
1.2 Discrete models: how to set up ODEs	17
1.3 Discrete forms	28
1.4 Differential forms	37
1.5 Solution sets of ODEs	43
1.6 Separation of variables in ODEs	61
1.7 The method of integrating factors	65
1.8 Change of variables in ODEs	67
1.9 Euler’s method: back to the discrete	73
1.10 How large is the difference between the discrete and the continuous?	79
1.11 Qualitative analysis of ODEs	88
1.12 Linearization of ODEs	96
1.13 Motion under forces: the acceleration	100
1.14 Discrete models: how to set up ODEs of second order	106
1.15 Discrete forms, continued	116
■ Chapter 2: Vector variables	126
2.1 Where matrices come from	126
2.2 Transformations of the plane	133
2.3 Linear operators	141
2.4 Examining and building linear operators	149
2.5 The determinant of a matrix	162
2.6 It’s a stretch: eigenvalues and eigenvectors	168
2.7 The significance of eigenvectors	179
2.8 Bases	184
2.9 Classification of linear operators according to their eigenvalues	191
■ Chapter 3: Vector and complex variables	198
3.1 Algebra of linear operators and matrices	198
3.2 Compositions of linear operators	204
3.3 How complex numbers emerge	209
3.4 Classification of quadratic polynomials	216
3.5 The complex plane \mathbf{C} is the Euclidean space \mathbf{R}^2	219
3.6 Multiplication of complex numbers: \mathbf{C} isn’t just \mathbf{R}^2	223
3.7 Complex functions	228
3.8 Complex linear operators	233
3.9 Linear operators with complex eigenvalues	236
3.10 Complex calculus	242
3.11 Series and power series	244
3.12 Solving ODEs with power series	250
■ Chapter 4: Systems of ODEs	254
4.1 Parametric curves	254
4.2 The predator-prey model	263

4.3 Qualitative analysis of the predator-prey model	267
4.4 Solving the Lotka–Volterra equations	270
4.5 Vector fields and systems of ODEs	273
4.6 Discrete systems of ODEs	279
4.7 Qualitative analysis of systems of ODEs	283
4.8 The vector notation and linear systems	287
4.9 Classification of linear systems	293
4.10 Classification of linear systems, continued	298
■ Chapter 5: Applications of ODEs	306
5.1 Vector-valued forms	306
5.2 The pursuit curves	307
5.3 ODEs of second order as systems	312
5.4 Vector ODEs of second order: a double spring	315
5.5 A pendulum	321
5.6 Planetary motion	324
5.7 The two- and three-body problems	330
5.8 A cannon is fired...	336
5.9 Boundary value problems	339
■ Chapter 6: Partial differential equations	343
6.1 Heat transfer between adjacent objects	343
6.2 Heat transfer depends on permeability	353
6.3 Heat transfer is caused by temperature differences	360
6.4 Heat transfer depends on the geometry	365
6.5 The heat PDE	370
6.6 Cells and forms in higher dimensions	375
6.7 Heat transfer in dimension 2: a plate	383
6.8 The heat PDE for dimension 2	391
6.9 Wave propagation in dimension 1: springs and strings	393
6.10 The wave PDE	400
6.11 Wave propagation in dimension 2: a membrane	405
Exercises	410
1 Exercises: Basics	410
2 Exercises: Analytical methods	412
3 Exercises: Euler’s method	413
4 Exercises: Generalities	414
5 Exercises: Models and setting up ODEs	415
6 Exercises: Qualitative analysis	416
7 Exercises: Systems	417
8 Exercises: Second order	418
9 Exercises: Advanced	419
10 Exercises: PDEs	420
11 Exercises: Computing	421
Index	422

Chapter 1: Ordinary differential equations

Contents

1.1 Incremental motion	12
1.2 Discrete models: how to set up ODEs	17
1.3 Discrete forms	28
1.4 Differential forms	37
1.5 Solution sets of ODEs	43
1.6 Separation of variables in ODEs	61
1.7 The method of integrating factors	65
1.8 Change of variables in ODEs	67
1.9 Euler’s method: back to the discrete	73
1.10 How large is the difference between the discrete and the continuous?	79
1.11 Qualitative analysis of ODEs	88
1.12 Linearization of ODEs	96
1.13 Motion under forces: the acceleration	100
1.14 Discrete models: how to set up ODEs of second order	106
1.15 Discrete forms, continued	116

1.1. Incremental motion

We can easily derive the speed from the distance that we have covered:

My speedometer is broken!
How can I know **how fast** I am going?!

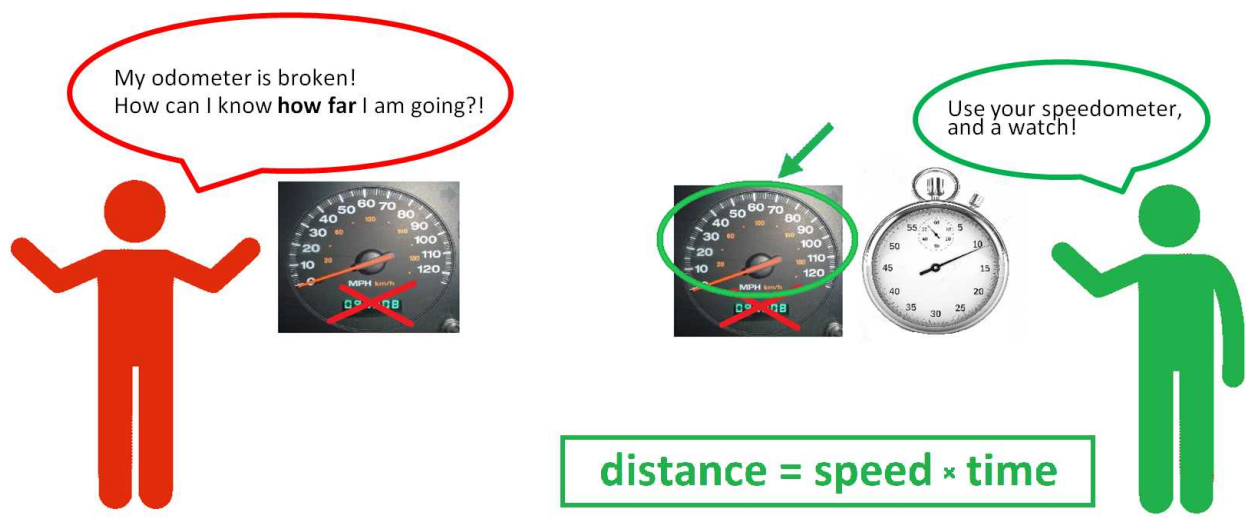


Use your odometer,
and a watch!



speed = distance / time

On the flip side, the derive the distance we have covered from the known velocity:



The two problems are solved, respectively, with the help of these two versions of the same elementary school formula:

speed = distance / time and distance = speed × time .

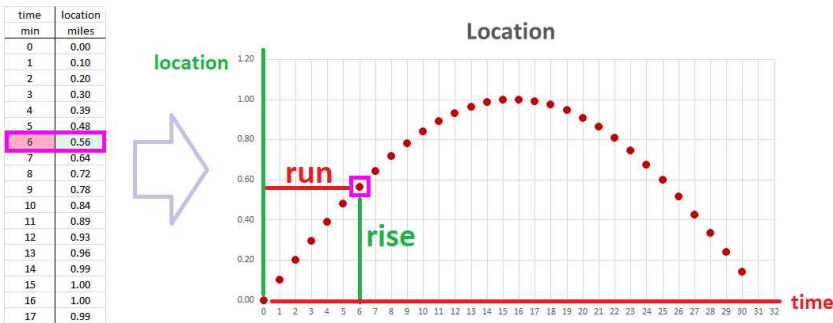
What takes this idea beyond elementary school is the possibility that *velocity varies*. The simplest case is when it varies incrementally.

We next consider more complex examples of the relation between location and velocity. First, *from location to velocity*...

Suppose that this time we have a *sequence* of more than 30 data points (more is indicated by "..."); they are the locations of a moving object recorded every minute:

time	minutes	0	1	2	3	4	5	6	7	8	9	10	...
location	miles	0.00	0.10	0.20	0.30	0.39	0.48	0.56	0.64	0.72	0.78	0.84	...

This data is also seen in the first two columns of the spreadsheet (left):



Every pair of numbers in the table is then plotted (right). To understand how fast we move over these one-minute intervals, we compute the *differences* of locations for each pair of consecutive locations.

First, the table.

We use the data from the row of locations. This is how the first one is computed:

time	min	0	1	...
location	miles	0.00	0.10	...
difference		↘	↓	...
			0.10 − 0.00	...
velocity	miles/min			...
			0.10	...

We compute this difference for each pair of consecutive locations and then place it in a row for the velocities

that we created at the bottom of our table:

time	min	0	1	2	3	4	5	6	7	8	9	...
location	miles	0.00	0.10	0.20	0.30	0.39	0.48	0.56	0.64	0.72	0.78	...
velocity	miles/min		↘ ↓	↘ ↓	↘ ↓	↘ ↓	↘ ↓	↘ ↓	↘ ↓	↘ ↓	↘ ↓	...
			0.10	0.10	0.10	0.09	0.09	0.09	0.08	0.07	0.07	...

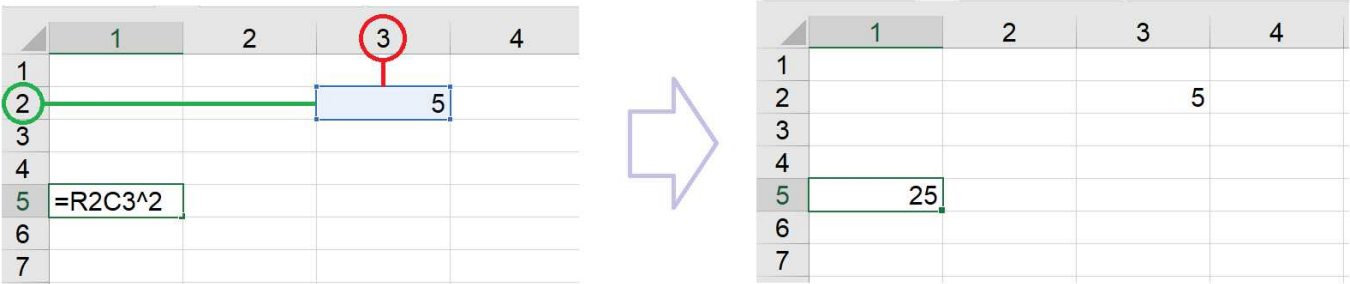
Practically, we'd rather use the computing capabilities of the spreadsheet.

Example 1.1.1: spreadsheet formulas

We use formulas to pull data from other cells. There are two ways. First, the “absolute” reference:

`=R2C3^2`

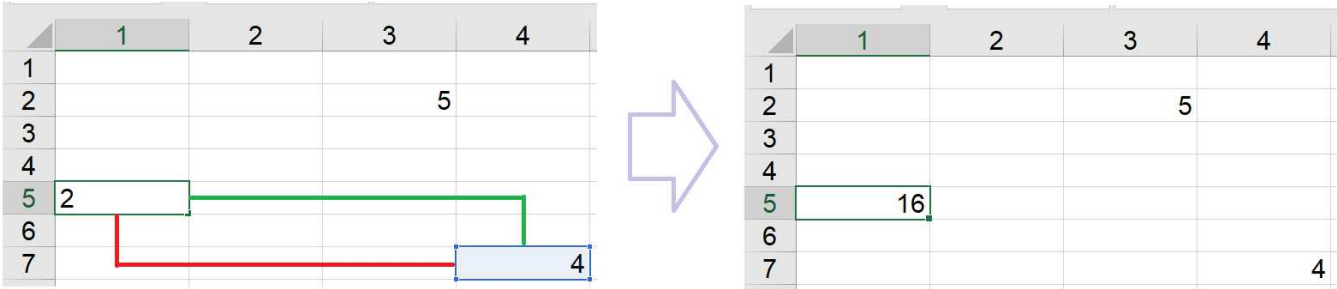
Any cell with this formula will take the value contained in the cell located at row 2 and column 3 and square it:



Second, the “relative” reference:

`=R[2]C[3]^2`

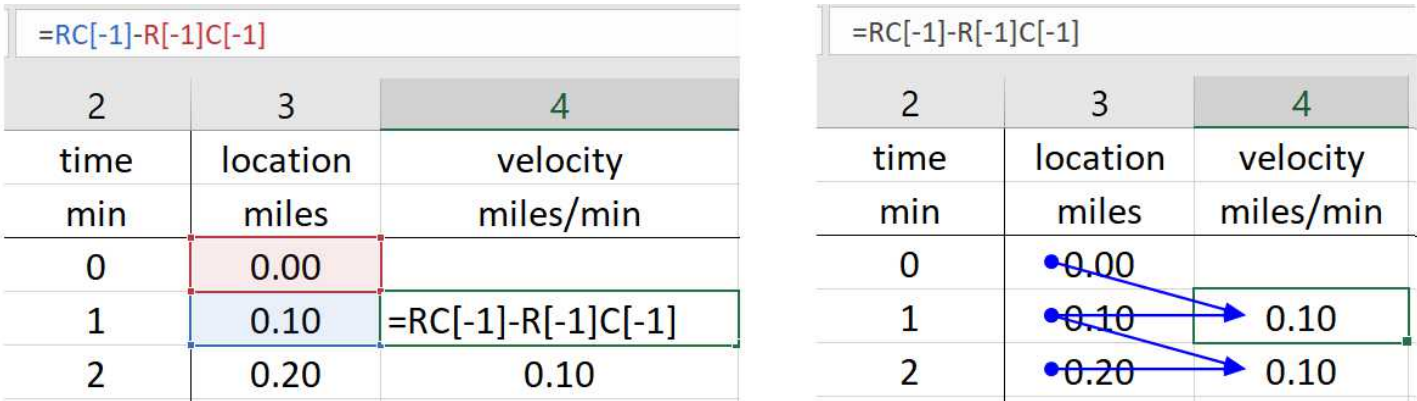
Any cell with this formula will take the value contained in the cell located 2 rows down and 3 columns right from it and square it:



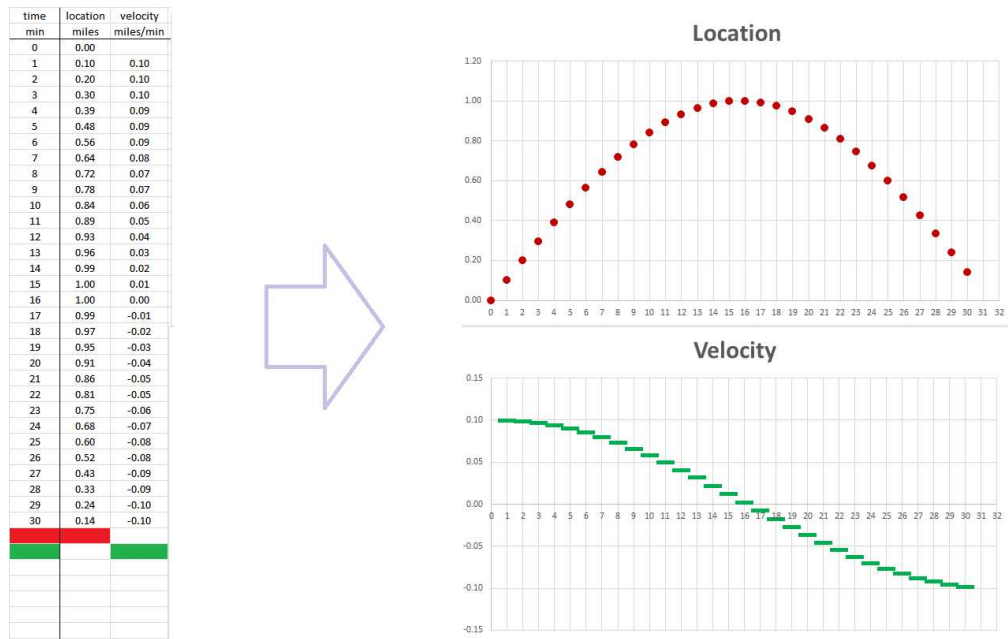
We compute the differences by pulling data from the column of locations with the following formula:

`=RC[-1]-R[-1]C[-1]`

Here, the two values come from the last column, `C[-1]`, same row, `R`, and last row, `R[-1]`. Below, you can see the two references in the formulas marked with red and blue (left) and the dependence shown with the arrows (right):



We place the result in a new column we created for the velocities:



This new data is illustrated with the second scatter plot. To emphasize the fact that the velocity data, unlike the location, is referring to time intervals rather than time instances, we plot it with horizontal segments. In fact, the data table can be rearranged as follows to make this point clearer:

time	0		1		2		3		4		...
location	0.00	—	0.10	—	0.20	—	0.30	—	.39	—	...
velocity	.	0.10	.	0.10	.	0.10	.	0.09	.	0.09	...

In a more general setting, the time progresses in increments. Let’s call it h . We can think of it as dependent on time if needed.

If a function f represents the locations at the time moments $a, a + h, a + 2h, \dots$, then the *displacement* over each period of time $[t, t + h]$ is the following expression:

$$\Delta f = f(t + h) - f(t)$$

This is a function is called the *difference* of f .

Furthermore, the *velocity* over each period of time $[t, t + h]$ is the following expression:

$$\frac{\Delta f}{\Delta t} = \frac{f(t + h) - f(t)}{h}$$

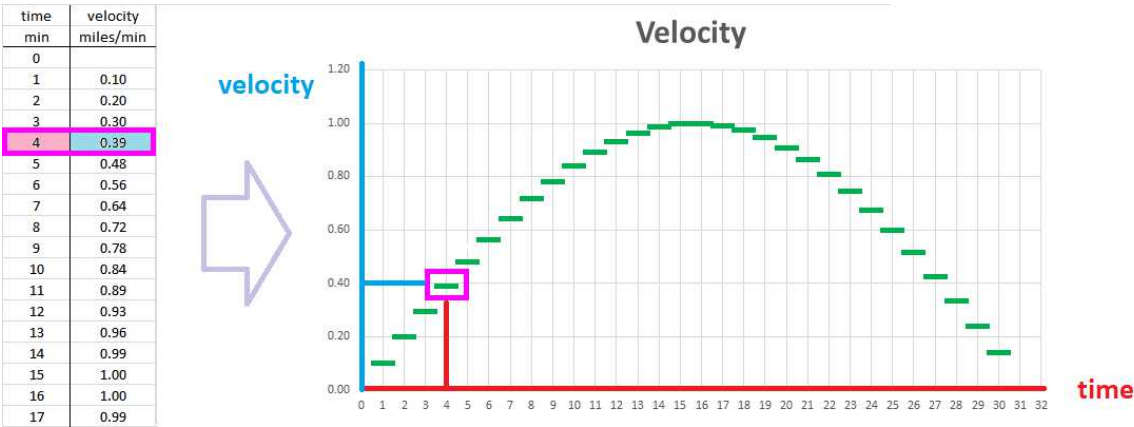
This is a function is called the *difference quotient* of f .

A more challenging and more important transition is *from velocity to location...*

Again, we consider 30 data points. These numbers are the values of the velocity of an object recorded every minute:

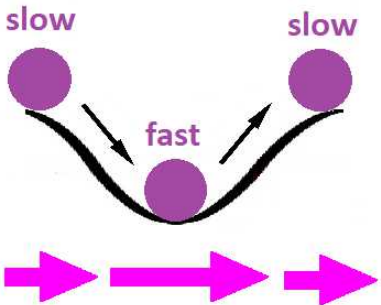
time	minutes	0	1	2	3	4	5	6	7	8	9	10	...
velocity	miles/hour		0.10	0.20	0.30	0.39	0.48	0.56	0.64	0.72	0.78	0.84	...

This data is also seen in the first two columns of the spreadsheet plotted one bar at a time:



This part data is furthermore illustrated as a scatter plot on the right. Again, we emphasize the fact that the velocity data is referring to time intervals by plotting its values with horizontal bars.

This particular set of data may be describing the horizontal speed of a ball rolling through a trough:



To find out where we are at the end of each of these one-minute intervals, we compute the *sum* of the consecutive velocities as the displacements for each interval by pulling the data from the row of velocities with the previous result. This is how the first one is computed, under the assumption that the initial location is 0:

time	min	0	1	...
velocity	miles		0.10	...
			↓	
sum		0.00+	0.10	...
		↑		
location	miles/min	0.00	0.10	...

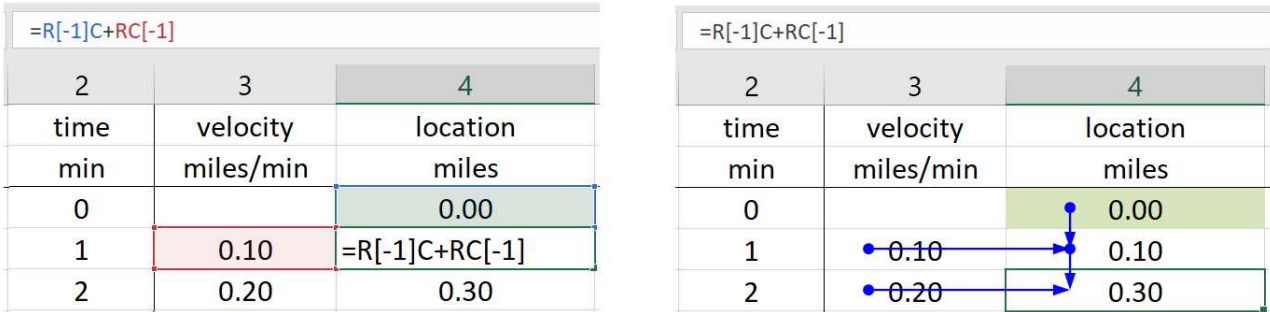
We place this data in a new row added to the bottom of our table:

time	min	0	1	2	3	4	5	6	7	8
velocity	miles		0.10	0.20	0.30	0.39	0.48	0.56	0.64	0.72
			↓	↓	↓	↓	↓	↓	↓	↓
location	miles/min	0.00 →	0.10 →	0.30 →	0.59 →	0.98 →	1.46 →	2.03 →	2.67 →	3.39 →

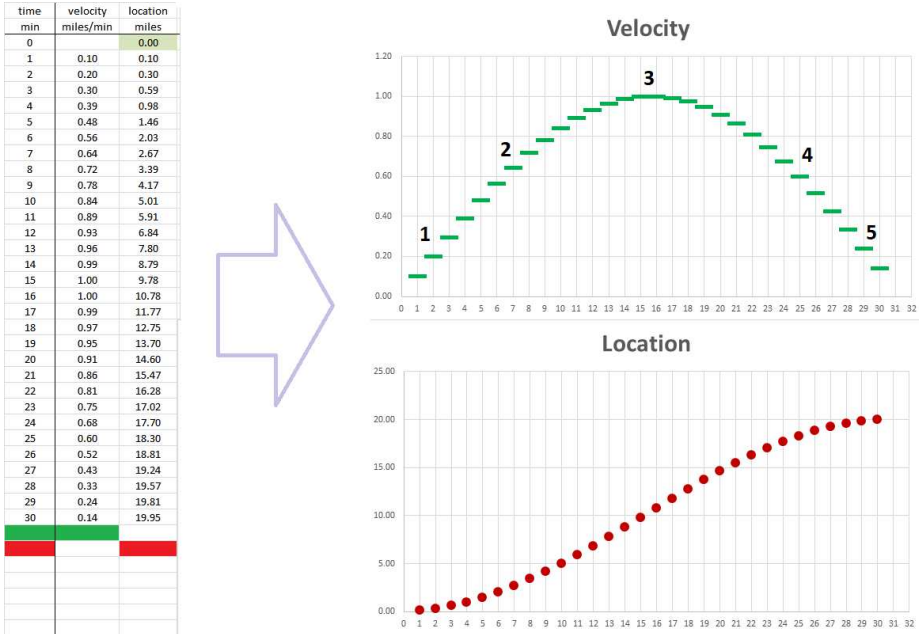
Practically, we use the spreadsheet. We compute the sums by pulling the data from the column of velocities using the following formula:

=R[-1]C+RC[-1]

Here, the two values come from the same, **C**, or last, **C[-1]**, column and the same, **R**, and last, **R[-1]**, row, as follows:



We place the result in a new column for the locations:



The data is also illustrated as the second scatter plot on the right.

We, again, rearrange the data table to make the difference between the two types of data clearer:

time	0		1		2		3		4		...
velocity	.	0.00	.	0.10	.	0.20	.	0.30	.	0.39	...
location	0.00	—	0.10	—	0.30	—	0.59	—	.98	—	...

In a more general setting, the time progresses in increments, h .

If a function g represents the displacements over each time interval $[t, t + h]$, then the *total displacement* over the period of time $[a, a + nh]$ is the following expression:

$$\sum_{k=1}^n g(a + kh).$$

This is a function is called the *sum* of g and it is computed recursively as explained above.

Furthermore, if a function v represents the velocities over each interval $[t, t + h]$, then the *total displacement* over the period of time $[a, a + nh]$ is the following expression:

$$\sum_{k=1}^n v(a + kh)h.$$

This is a function is called the *Riemann sum* of v .

In the examples above, the values of g and v are placed at the *ends* of the intervals. Where else? In the general setup presented below, we transcend this issue.

1.2. Discrete models: how to set up ODEs

Below, we produce discrete models and then differential equations from verbal descriptions.

We have a *partition* of an interval $[a, b]$ or $[a, +\infty]$. It is, first, a sequence of n *nodes*, t_i :

$$a = t_0 < t_1 < t_2 < \dots$$

The *increments* of t are the lengths of the intervals:

$$\Delta t_i = t_i - t_{i-1}, \ i = 1, 2, \dots$$

In addition to the nodes, there are *edges*:

$$c_1 = [t_0, t_1], \ c_2 = [t_1, t_2], \ \dots$$

We then speak of a *cell decomposition* of the interval.

In all examples in this section, we assume that the increments are equal:

$$\Delta t_i = t_i - t_{i-1} = h, \ i = 1, 2, \dots$$

Therefore, we have:

$$a = t_0, \ t_1 = a + h, \ t_2 = a + 2h, \ \dots$$

Next, the *difference* of a function y defined at the primary nodes is a function defined at the edges of the partition. On the edge $[t_{k-1}, t_k]$, we have:

$$\Delta y = y(t_k) - y(t_{k-1}), \ k = 1, 2, \dots$$

Suppose the initial value $y(a)$ is set and the difference of y is given one step at a time. Then we find y by a recursive formula:

$$y(t_{k+1}) = y(t_k) + \Delta y, \ k = 0, 1, 2, \dots$$

Meanwhile, this is the familiar *difference quotient*:

$$\frac{\Delta y}{\Delta t} = f.$$

Example 1.2.1: uniform motion

Uniform motion means that we cover the same distance over equal intervals of time. If y is our location, this translates into: The increment of the location Δy is proportional to the increment of time Δt . The algebra is as follows:

$$\Delta y = k \cdot \Delta t,$$

for some constant k .

If we take into account a familiar idea from physics, we have a *description of the dynamics*:

► “The velocity is constant”.

The algebra is as follows:

$$\frac{\Delta y}{\Delta t} = k.$$

Suppose we also have an initial condition:

$$y(t_0) = y_0.$$

Then our equation gives us a sequence via this *recursive formula*:

$$y(t_{n+1}) = y(t_n) + k \cdot \Delta t.$$

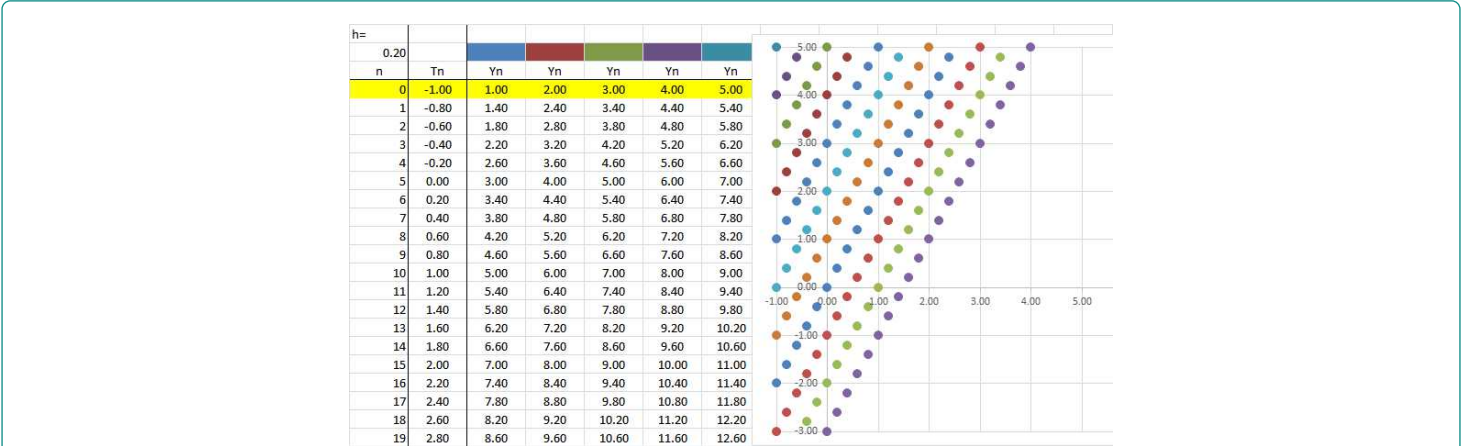
As an illustration, consider a “difference equation”:

$$\frac{\Delta y}{\Delta t} = 2.$$

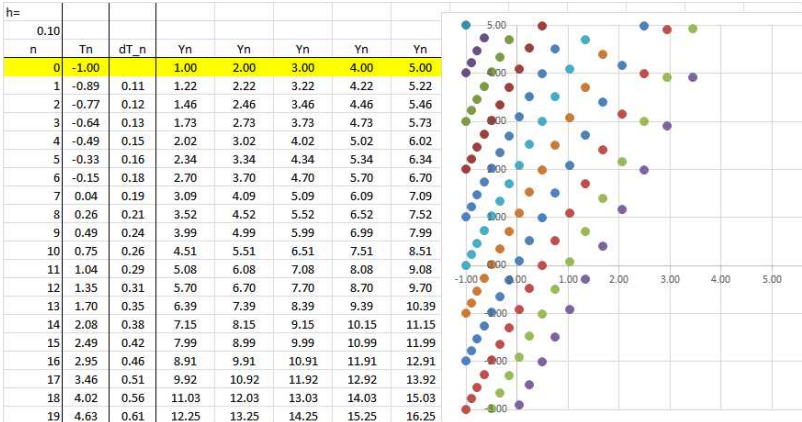
We choose the time increments to be constant $\Delta t = h = 0.2$. Then:

$$y(t_{n+1}) = y(t_n) + 2h.$$

Here are a few solutions for various initial conditions (on the y -axis):



They are just arithmetic progressions: step right by h and then up by $2h$. Generally, they don't have to be – when the partition is uneven:



But the points still lie on the same straight lines!

Example 1.2.2: location from velocity

The formula has been used in Volumes 2, 3, and 4 to model uniform motion and acquire location from this velocity. More generally, we have:

- “The velocity depends on time”.

We take any function $z = f(t)$ defined at the edges of the decomposition and think of $f(c_i)$ as the value of the velocity during the time interval $[t_{i-1}, t_i]$. We set up a difference equation:

$$\frac{\Delta y}{\Delta t} = f(c_n).$$

Rearranged it produces its own solution as a recursive formula:

$$y(t_{n+1}) = y(t_n) + f(c_n) \cdot \Delta t,$$

where

$$t_{n+1} = t_n + \Delta t.$$

What kind of function is f ? It is function of one variable $z = f(t)$ (plotted below left). However, the graph of a solution $y_n = y(t_n)$ is to be plotted on the ty -plane. It is, therefore, beneficial to think of f as a function of two variables $z = f(t, y)$ that just happens to have the same value for each t regardless of y . We sample f over its domain, a rectangle of t 's and y 's, and write the outputs (middle):

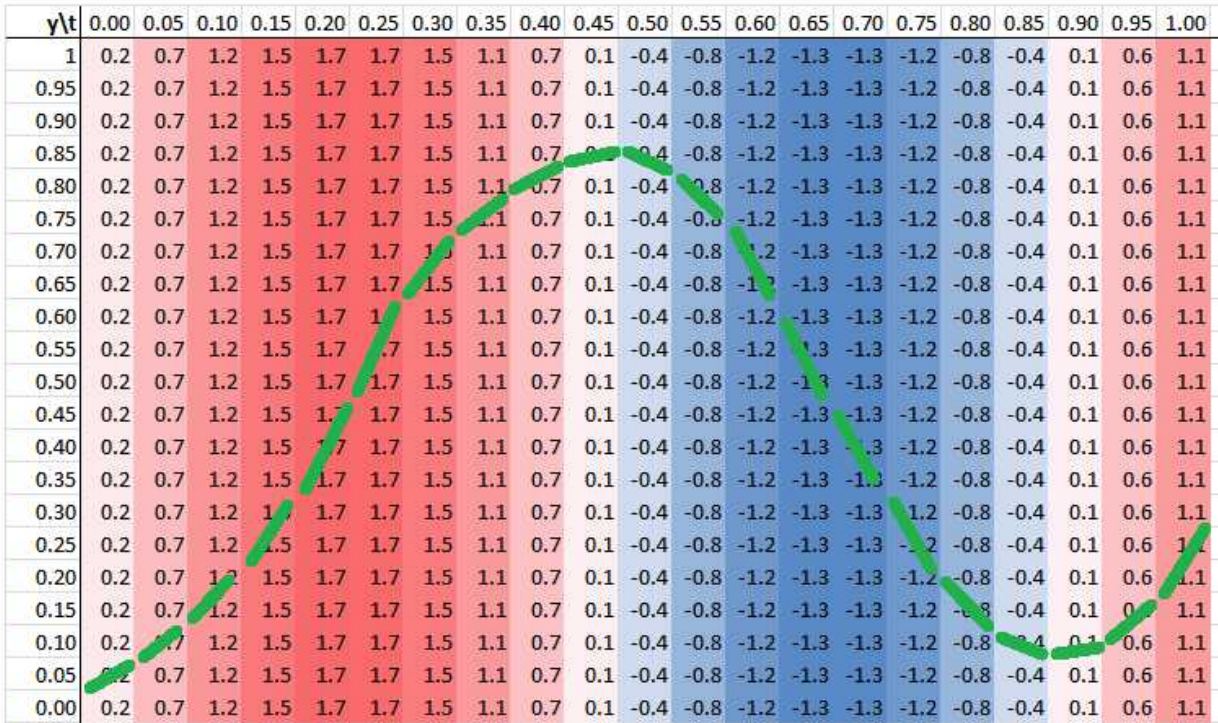


We enhance the visualization by showing the values of f in terms of color – blue for negative and red for positive. We also show f as a surface (right). The middle is the most convenient of these as it is seen as a *field of velocities*.

So, as y is plotted point by point, we know the procedure for constructing a solution:

- We go up when the area is red.
- We go down when the area is blue.

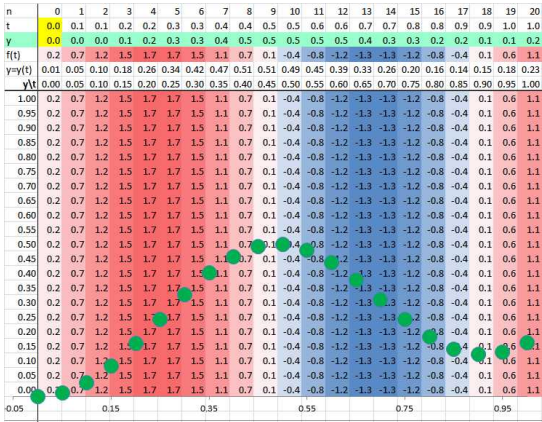
We can draw a curve in this manner:



More precisely, we read the data from the table as we progress one point at a time:

velocity		time interval h		jump Δy
0.2	.	0.1	=	0.02
0.7	.	0.1	=	0.07
1.2	.	0.1	=	0.12
..	

Or it is done by the computer:



In the meantime, we go back to our difference equation:

$$\frac{\Delta y}{\Delta t} = f.$$

Furthermore, what if f isn't just a stream of numbers but a sampled continuous function? And what if y is a sampled continuously differentiable function? We can take the limit of the latter equation,

$\Delta t \rightarrow 0$. The result is a differential equation:

$$\frac{dy}{dt} = f,$$

or

$$y' = f(t).$$

Exercise 1.2.3

Plot more solutions starting at other points. What is a relation between them?

Example 1.2.4: data stream

The quantities used to compute the next location (or a state) – current time and velocity – doesn’t have to come from a formula! This is nothing but data and it may come as *strings of numbers* (or it can be fed into the computer continuously).

For example, the initial time t_0 and the initial location y_0 are placed in the first row of the spreadsheet and, as we progress in time and space, new numbers are placed in the next row of our spreadsheet:

$$t_n, \, v_n, \, y_n, \, n = 1, 2, 3, \dots$$

The same recursive formula is used to find the next location:

$$y_{n+1} = y_n + v_{n+1} \cdot \Delta t.$$

The result is a growing table of values:

	iteration n	time t_n	velocity v_n	location y_n
initial:	0	3.5	--	22
	1	3.6	33	25.3

	1000	103.5	4	336

All but the last column may come as a data file.

Example 1.2.5: exponential growth

The dynamics is very different in the next description:

► “The rate of growth/decay of the population is proportional to the current population.”

An example is bacteria or rabbits growing with an unrestricted amount of food because the number of the newborn is proportional to the number of adults.

We rewrite the description:

$$\frac{\Delta y}{\Delta t} = ky,$$

for some constant k . In other words, the increment of the population Δy is proportional to y as well as Δt :

$$\Delta y = k \cdot y \cdot \Delta t.$$

The “right-hand side” function is very simple:

$$f(y) = ky.$$

Once again, however, the graph of a solution $y_n = y(t_n)$ is plotted on the ty -plane and, therefore, it is better to think of f as a function of two variables $z = f(t, y)$ that just happens to have the same value for each y regardless of t :

Exercise 1.2.7

Plot solutions for $k < 0$. What does the system model?

Example 1.2.8: logistic growth

A different, and more complex, population dynamics is give by the following description:

- “The rate of growth of the population is proportional to the current population and to the remaining population of the total that can be sustained”.

An example is bacteria growing in a jar.

We rewrite the description:

$$\frac{\Delta y}{\Delta t} = ky(T - y) ,$$

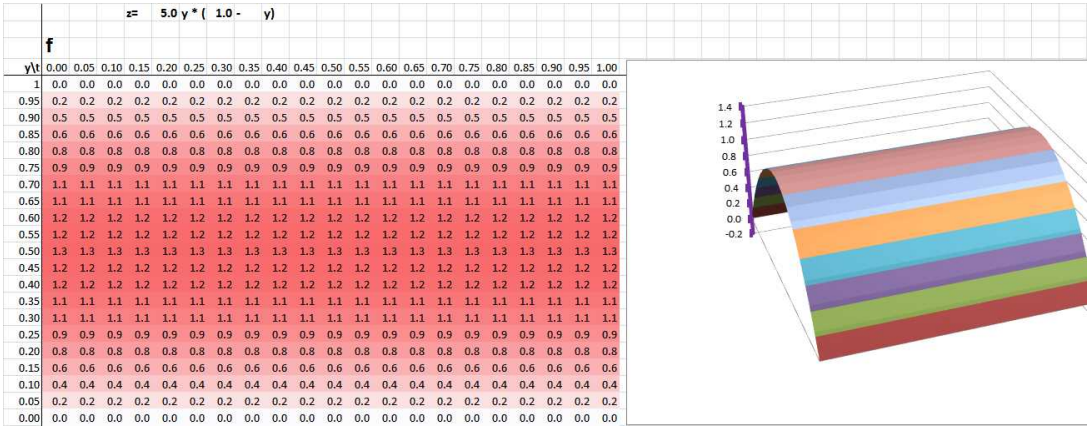
for some constant k . In other words, the increment of the population Δy is proportional to y ... and $T - y$, where T is the total possible population (as well as Δt as before):

$$\Delta y = k \cdot y \cdot (T - y) \cdot \Delta t .$$

The “right-hand side” function is slightly more complex:

$$f(t,y) = ky(T - y) .$$

This is our function ($k = 5, T = 1$):



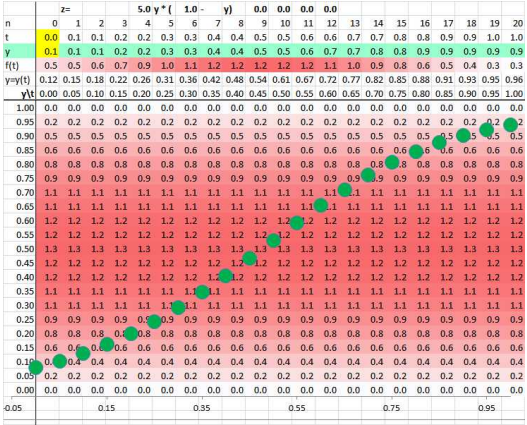
We see low level of potential growth in the areas where the population is low or where it is too close to be exhausting the resources. With an initial condition:

$$y(t_0) = y_0 ,$$

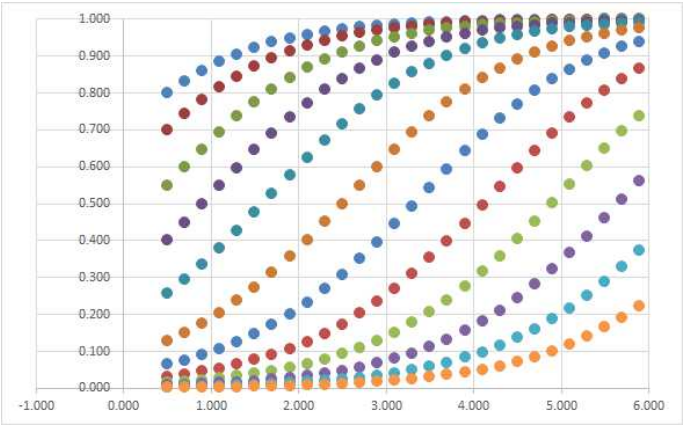
we produce a discrete model via this recursive formula:

$$y(t_{n+1}) = y(t_n) + ky(t_n)(T - y(t_n)) \cdot \Delta t, \quad k > 0 .$$

The model describes a restricted growth ($k = 5, \Delta t = h = .05$):



We see the fastest growth when the population is about half of what can be sustained. All solutions exhibit this behavior:



Choosing Δt small relative to k could break our model ($k = 25, \Delta t = .05$):

	50.0 y * (1.0 - y)																					
n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
t	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1.00	
y	0.10	0.33	0.87	1.15	0.72	1.22	0.54	1.16	0.70	1.22	0.54	1.16	0.70	1.22	0.54	1.16	0.70	1.22	0.54	1.16	0.70	
f(y,t)	4.50	5.53	5.53	5.53	5.53	5.53	5.53	5.53	5.53	5.53	5.53	5.53	5.53	5.53	5.53	5.53	5.53	5.53	5.53	5.53	5.53	
y=y(t)	0.33	0.87	1.15	0.72	1.22	0.54	1.16	0.70	1.22	0.54	1.16	0.70	1.22	0.54	1.16	0.70	1.22	0.54	1.16	0.70	1.22	
y(t)	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1.00	
	1.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
	0.95	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	
	0.90	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	
	0.85	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	
	0.80	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	
	0.75	9.4	9.4	9.4	9.4	9.4	9.4	9.4	9.4	9.4	9.4	9.4	9.4	9.4	9.4	9.4	9.4	9.4	9.4	9.4	9.4	
	0.70	10.5	10.5	10.5	10.5	10.5	10.5	10.5	10.5	10.5	10.5	10.5	10.5	10.5	10.5	10.5	10.5	10.5	10.5	10.5	10.5	
	0.65	11.4	11.4	11.4	11.4	11.4	11.4	11.4	11.4	11.4	11.4	11.4	11.4	11.4	11.4	11.4	11.4	11.4	11.4	11.4	11.4	
	0.60	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	
	0.55	12.4	12.4	12.4	12.4	12.4	12.4	12.4	12.4	12.4	12.4	12.4	12.4	12.4	12.4	12.4	12.4	12.4	12.4	12.4	12.4	
	0.50	12.5	12.5	12.5	12.5	12.5	12.5	12.5	12.5	12.5	12.5	12.5	12.5	12.5	12.5	12.5	12.5	12.5	12.5	12.5	12.5	
	0.45	12.4	12.4	12.4	12.4	12.4	12.4	12.4	12.4	12.4	12.4	12.4	12.4	12.4	12.4	12.4	12.4	12.4	12.4	12.4	12.4	
	0.40	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	
	0.35	11.4	11.4	11.4	11.4	11.4	11.4	11.4	11.4	11.4	11.4	11.4	11.4	11.4	11.4	11.4	11.4	11.4	11.4	11.4	11.4	
	0.30	10.5	10.5	10.5	10.5	10.5	10.5	10.5	10.5	10.5	10.5	10.5	10.5	10.5	10.5	10.5	10.5	10.5	10.5	10.5	10.5	
	0.25	9.4	9.4	9.4	9.4	9.4	9.4	9.4	9.4	9.4	9.4	9.4	9.4	9.4	9.4	9.4	9.4	9.4	9.4	9.4	9.4	
	0.20	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	8.0	
	0.15	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	
	0.10	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	
	0.05	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4	
	0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

The population is sometimes larger than the capacity of the jar!

If y is a sampled differentiable function, then the limit of the difference equation

$$\frac{\Delta y}{\Delta t} = y(T - y) \, .$$

produces the following differential equation:

$$y' = ky(T - y) \, .$$

Exercise 1.2.9

What does the case of $k < 0$ model?

Example 1.2.10: Newton’s Law of Cooling

A familiar phenomenon has the following dynamics:

- “The rate of cooling is proportional to the difference of the current temperature and the room temperature”.

An example is cooling of a cup of coffee or warming up a can of soda.

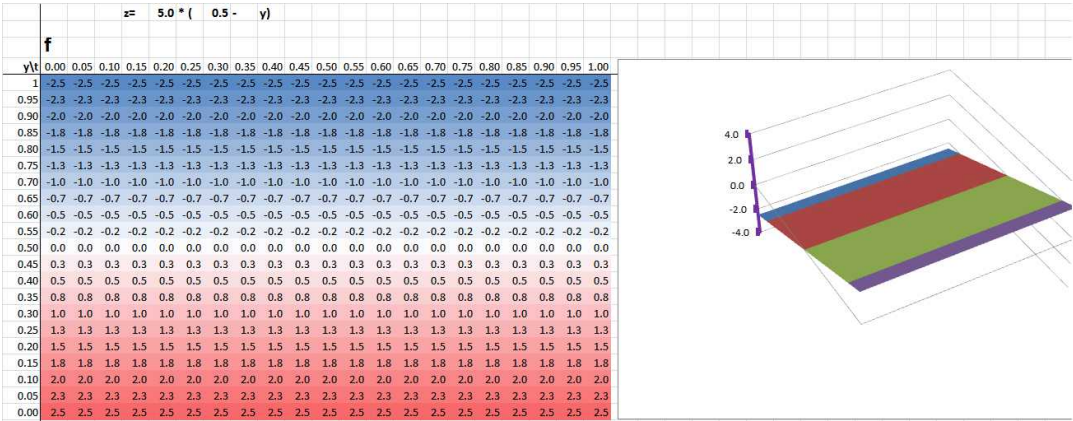
We rewrite the description:

$$\frac{\Delta y}{\Delta t} = k(r - y) \, ,$$

where r is the room temperature, for some constant $k > 0$. The “right-hand side” function is slightly simpler than last:

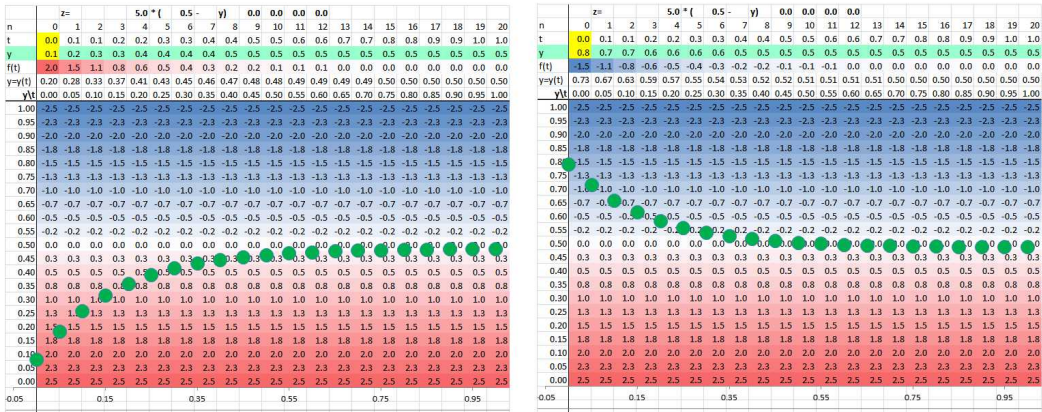
$$f(t, y) = k \cdot (r - y) \, .$$

It is plotted below ($k = 10, \Delta t = .05$):

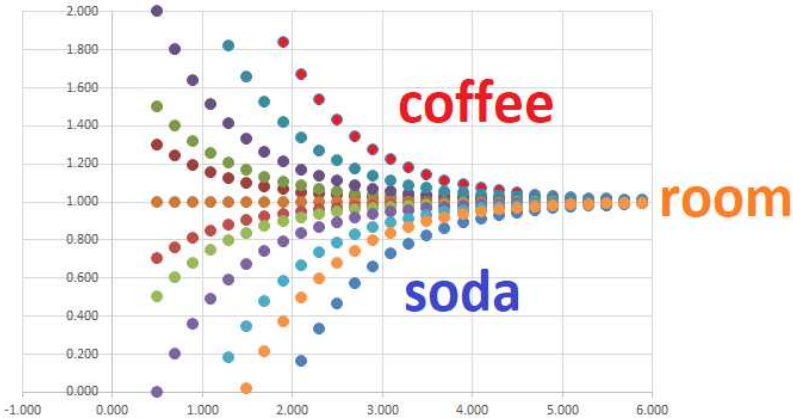


We can see that when the temperature is higher than the room temperature, it decreases and when the temperature is lower than the room temperature, it increases. This is why the object’s temperature isn’t expected to “overshoot” that of the room. The equation gives us a solution via this recursive formula:

$$y(t_{n+1}) = y(t_n) + k(r - y(t_n)) \cdot \Delta t.$$



The formula was used in [Chapter 2DC-6](#) to model Newton’s Law of Cooling:



In the meantime, the derivative, if any, would satisfy the following:

$$y' = k(r - y).$$

Example 1.2.11: Newton’s Law of Cooling continued

Choosing Δt small relative to k could break our model ($k = 25, h = \Delta t = .05$):

location:

$$v_n = f(y_n).$$

But the velocity of the flow may vary with time! We have then:

$$v_n = f(t_n, y_n).$$

Once again, we think of y_n as a function: $y_n = y(t_n)$. We substitute this, as well as $t = t_n$, into the recursive formula for y_{n+1} :

$$y(t + \Delta t) = y(t) + f(t, y(t)) \cdot \Delta t,$$

or

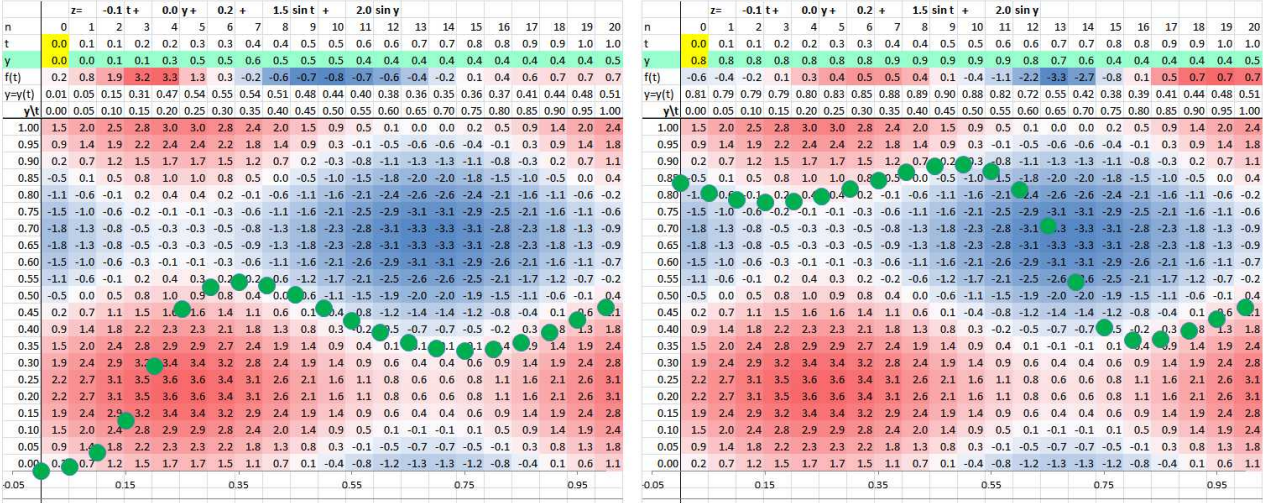
$$\Delta y = f(t, y, t) \cdot \Delta t.$$

Furthermore,

$$\frac{\Delta y}{\Delta t} = f(t, y).$$

These equations are the most general form of a discrete differential equation. We will use the former for simulation but will not attempt to *solve* the latter. In the rest of this chapter, we will try to solve some of the ODEs for derivatives that emerge from these relations: $y' = f(t, y)$.

This is an example of a possible choice of f and some solutions of the corresponding difference equation:



In the previous example we also derived locations from velocities; however, the velocities were dependent on *time*, while there they depends on *location*!

In summary, we follow the same path:

1. There is a quantity y – a number that represents the current state of the system – that may change with time t .
2. We know that the increment Δy of y over a standard interval of time is proportional to its length Δt .
3. We know that the increment Δy is proportional to another quantity f that depends on the current time t and the current state y .
4. We then form a difference equation (discrete ODE):

$$\Delta y = f(t, y) \Delta t.$$

5. We solve this equation recursively:

$$y(t + \Delta t) = y(t) + f(t, y(t)) \cdot \Delta t.$$

6. If this is a continuous phenomenon being sampled, we use the difference equation to form an equation

for the difference quotient:

$$\frac{\Delta y}{\Delta t} = f(t,y) .$$

7. Under the limit $\Delta t \rightarrow 0$, we obtain a differential equation:

$$\frac{dy}{dt} = f(t,y) .$$

8. We solve this equation (when possible) by finding the differentiable functions $y = y(t)$ that satisfy this equation.
9. When we are unable to solve the differential equation, we go back to the difference equation and its recursive solution (step 5). Beware of crude discretizations!

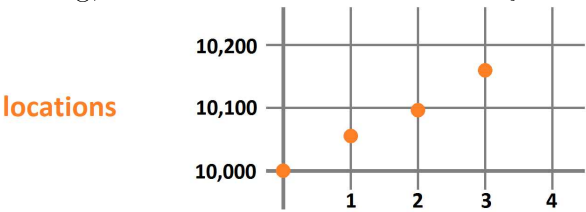
A simple example of such a failure is $f(t,y) = e^{-t^2}$.

1.3. Discrete forms

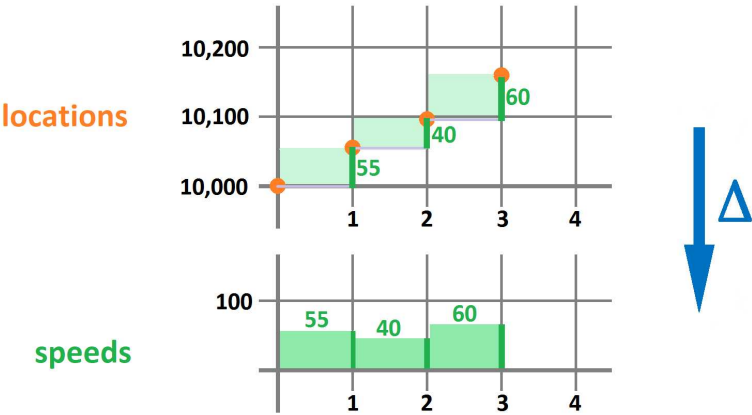
Some numerical simulation of motion that we have seen can be computed with tools similar to but simpler than functions defined at the nodes or the edges of a partition/decomposition.

Example 1.3.1: speedometer is broken

Recall what we started with in the very beginning. Suppose the speedometer is broken and in order to estimate how fast we are driving, we look at the odometer every hour:



That’s a discrete 0-form. To find the displacement for every hour we just look at the differences:



That’s a discrete 1-form. Alternatively, the odometer is broken and we look at the speedometer to sample the velocity and then, via the Riemann sums, find the displacement.

Let’s start over.

Suppose we have a partition of some interval $[a,b]$ in the x -axis or the whole axis.

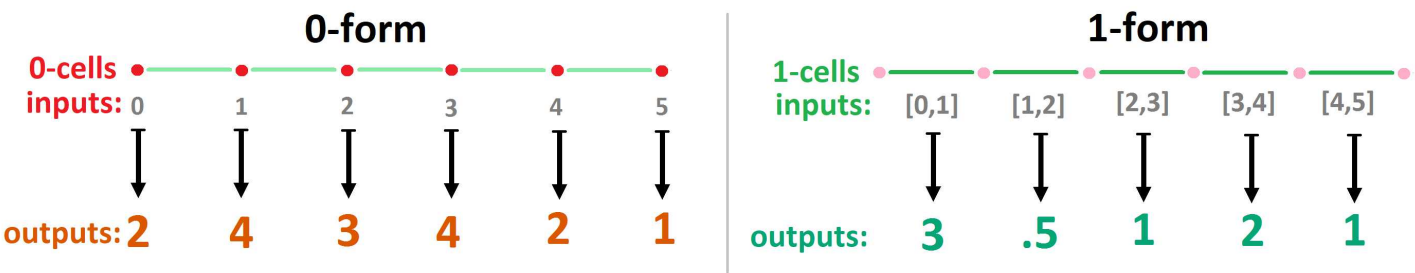
Definition 1.3.5: discrete forms

Suppose we have a cell decomposition (of an interval or the whole real line). Then *discrete forms* are defined as follows:

- A 0-form is a real-valued function with 0-cells (nodes) as inputs.
- A 1-form is a real-valued function with 1-cells (edges) as inputs.

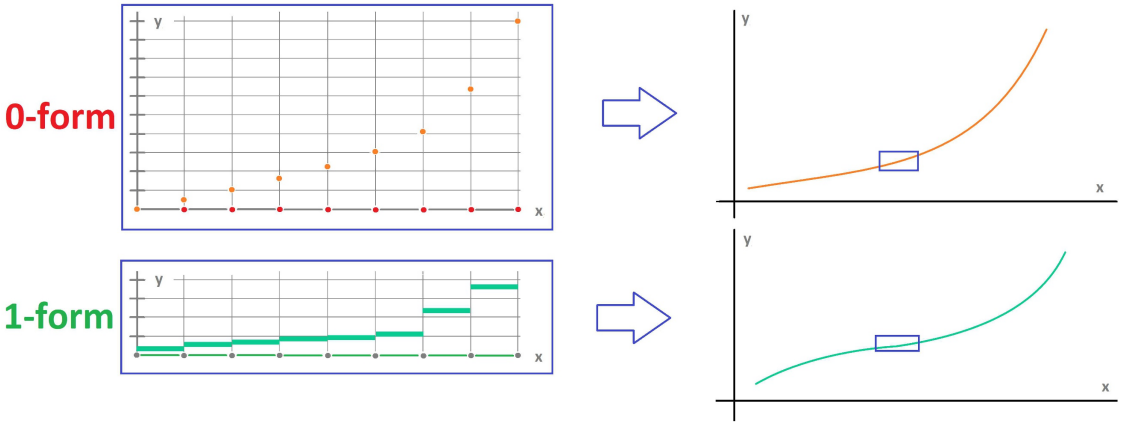
They are also called *forms of degree 0* and *1* respectively.

We use arrows to picture these functions as correspondences:



For a 0-form, x is a node, a number, and $y = f(x)$ is also a number. Together, they produce (x, y) , a point on the xy -plane (with the x -axis split into cells as shown above). For a discrete 1-form, $[A, B]$ is an interval in the x -axis, and $y = g([A, B])$ is a number. Together, they produce a collection of points on the xy -plane such as (x, y) for every x in $[A, B]$. The result is a horizontal segment.

Even though these functions may consist of unrelated pieces, it is possible that we can see a *continuous curve* if we zoom out:



Conversely, a continuous function is *sampled* in order to produce discrete forms.

Example 1.3.7: motion

Let’s consider an example of motion. Suppose a 0-form p gives the position of a person and suppose

- at time n hours we are at the 5 mile mark: $p(n) = 5$, and then
- at time $n + 1$ hours we are at the 7 mile mark: $p(n + 1) = 7$.

We don’t know what exactly has happened during this hour but the simplest assumption would be that we have been walking at a constant speed of 2 miles per hour.

5

n

7

n+1

2 = 7 - 5

2

n

n+1

Now, instead of our velocity function v assigning this value to each instant of time during this period, it is assigned to the *whole* interval:

$$v \Big|_{[n,n+1]} = 2 \text{ , or better } v\Big([n,n+1]\Big) = 2$$

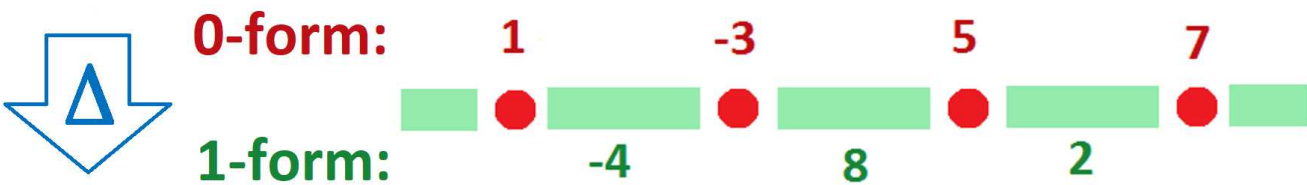
This way, the elements of the domain of the velocity function are the edges and the resulting function is a discrete 1-form!

The functions, when defined on the nodes change abruptly and, consequently, the change over every interval $[A, B]$ is simply the *difference of values* at the nodes, from right to left:

f(B) - f(A).

The output of this simple computation is then assigned to the interval $[A, B]$:

[A, B] ↦ f(B) - f(A).



Just as before, the difference stands for the change of the function.

Definition 1.3.8: difference

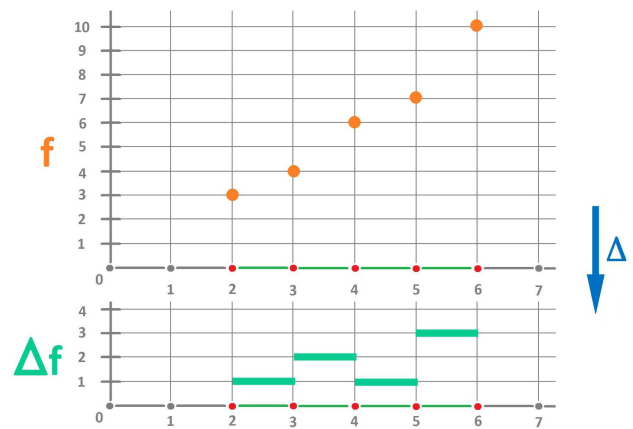
Suppose we have a cell decomposition of an interval with these edges:

$$a_k = [x_{k-1}, x_k] .$$

Then the *difference of a discrete 0-form* f is a discrete 1-form denoted by Δf and given by its values at k th edge a_k :

$(\Delta f) (a_k) = f(x_k) - f(x_{k-1})$

The relation between a 0-form and its difference is illustrated below:



Example 1.3.9: spreadsheet

This is how a spreadsheet computes the difference of a function given by the data in the first column:

Coordinates:	-0.4		-0.2		0.0		0.2		0.4		0.6	
Cells:												
Nodes:	-0.4		-0.2		0		0.2		0.4		0.6	
Edges:		[-.4,-.2]		[-.2,.0]		[.0,.2]		[.2,.4]		[.4,.6]		[.6,.8]
0-form	1		0.3		0.2		0.1		0		0	
difference		-0.7		-0.1		-0.1		-0.1		0		0.2

Example 1.3.10: algebra

When the discrete 0-forms are represented by formulas, the computations are straightforward ($h = 1$) with a chance of simplification:

- (1) $f(n) = 3n^2 + 1 \implies \Delta f(a_n) = (3n^2 + 1) - (3(n - 1)^2 + 1) = 6n - 3$
- (2) $g(n) = \frac{1}{n} \implies \Delta g(a_n) = \frac{1}{n} - \frac{1}{n - 1} = -\frac{1}{n(n - 1)}$ for $n \neq 0, 1$
- (3) $p(n) = 2^n \implies \Delta p(a_n) = 2^n - 2^{n-1} = 2^{n-1}$

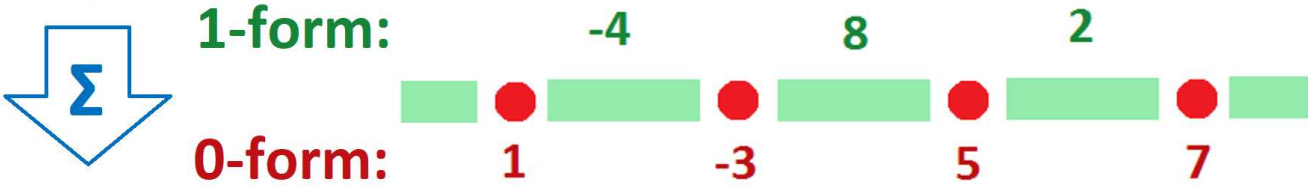
Example 1.3.11: derivatives

This is how we get from differences to difference quotients to derivatives. Let's take $f(x) = x^3$:

$\frac{\Delta f}{\Delta x}(x) = \frac{f(x+h) - f(x)}{h}$	The difference quotient is written from the definition.
$= \frac{(x+h)^3 - x^3}{h}$	The function is specified.
$= \frac{x^3 + 3x^2h + 3xh^2 + h^3 - x^3}{h}$	The numerator is expanded.
$= \frac{3x^2h + 3xh^2 + h^3}{h}$	The terms without h are cancelled.
$= 3x^2 + 3xh + h^2$	The numerator is divided by h .
$\frac{\Delta f}{\Delta x}(x) = 3x^2 + 3xh + h^2$	This is the simplified difference quotient.
as $h \rightarrow 0 \rightarrow 3x^2 + 3x \cdot 0 + 0^2$	The limit is then evaluated by substitution $h = 0$...
$= 3x^2$	The difference quotient is continuous with respect to h .

As you can see, without the limit, the time increment $\Delta t = h$ will remain a parameter of the model!

Now, the other fundamental construction of calculus.



Definition 1.3.12: sum

Suppose we have a cell decomposition of an interval with these edges:

$$a_k = [x_{k-1}, x_k] .$$

Then the *sum of a discrete 1-form g* is a discrete 0-form denoted by $\sum g$ and given by its value at each node x_k :

$$\left(\sum g\right) (x_k) = \sum_{i=1}^k g(a_i) = g(a_1) + g(a_2) + \dots + g(a_k)$$

Example 1.3.13: Riemann sums

How does this relate to the Riemann sum as we know it? Recall first that an *augmented partition* of an interval $[a, b]$ is a sequence of n *primary nodes*, x_i , alternating with the *secondary nodes*, c_i , of the partition:

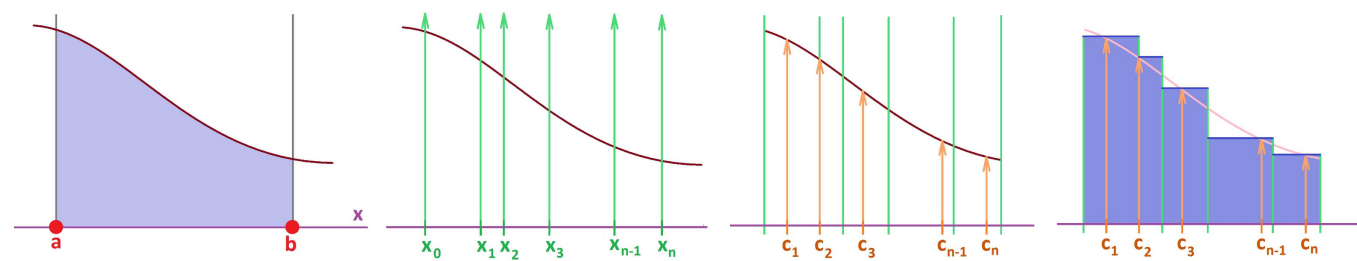
$$a = x_0 \leq c_1 \leq x_1 \leq c_2 \leq x_2 \leq \dots \leq x_{n-1} \leq c_n \leq x_n = b .$$

The *increments* of t are:

$$\Delta x_i = x_i - x_{i-1}, \quad i = 1, 2, \dots, n .$$

Typically, we have the intervals of equal length, $\Delta x_i = \Delta x = h$.

Suppose we have a function f defined on $[a, b]$. We then sample f at the secondary nodes:



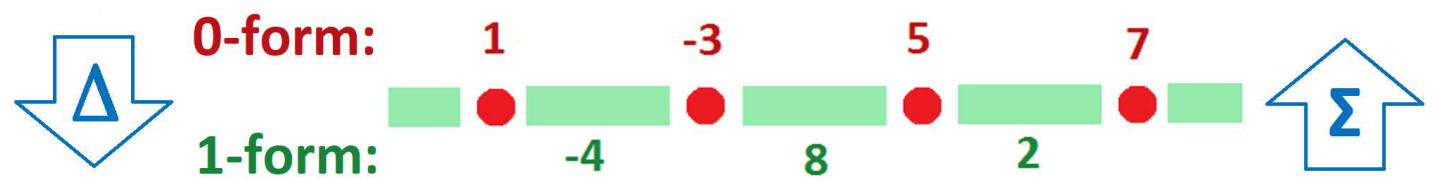
We then create a 1-form by specifying its value at the i th edge a_i :

$$g(a_i) = f(c_i)\Delta x_i.$$

Then the Riemann sum of f over $[a, b]$ (the total area of the bars) is just the sum of this special 1-form:

$$\sum_{i=1}^k f(c_i)\Delta x_i = f(c_1)\Delta x_1 + f(c_2)\Delta x_2 + \dots + f(c_k)\Delta x_k.$$

Next, *the fundamental relation between differences and sums.*



Suppose again we have a cell decomposition of an interval with these nodes and these edges:

$$a_k = [x_{k-1}, x_k].$$

First, we have a 0-form and a 1-form:

- If f is defined at the nodes x_k , $k = 0, 1, 2, \dots, n$, then
- the difference g of f is defined at the edges according to:

$$g(a_k) = f(x_k) - f(x_{k-1}).$$

Theorem 1.3.14: Fundamental Theorem of Discrete Calculus I

Suppose f is a discrete 0-form. Then, for each node x of the cell decomposition, we have:

$$\sum (\Delta f)(x) = f(x) - f(a)$$

Second, we have a 1-form and a 0-form:

- If g is defined at the edges a_k , $k = 1, 2, \dots, n$, then
- the sum f of g is defined recursively at the nodes according to:

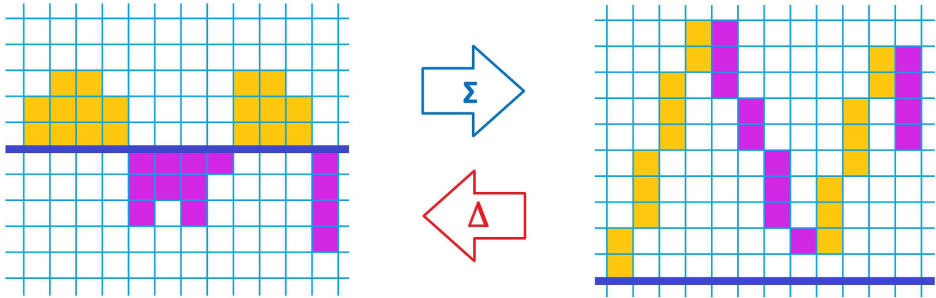
$$f(x_k) = f(x_{k-1}) + g(a_k).$$

Theorem 1.3.15: Fundamental Theorem of Discrete Calculus II

Suppose g is a discrete 1-form. Then, for each node x of the decomposition, we have:

$$\Delta \left(\sum g \right) ([a, x]) = g ([a, x])$$

So, the two operations cancel each other in either order:



Example 1.3.16: fundamental theorems, computed

We use a spreadsheet. Recall the formulas:

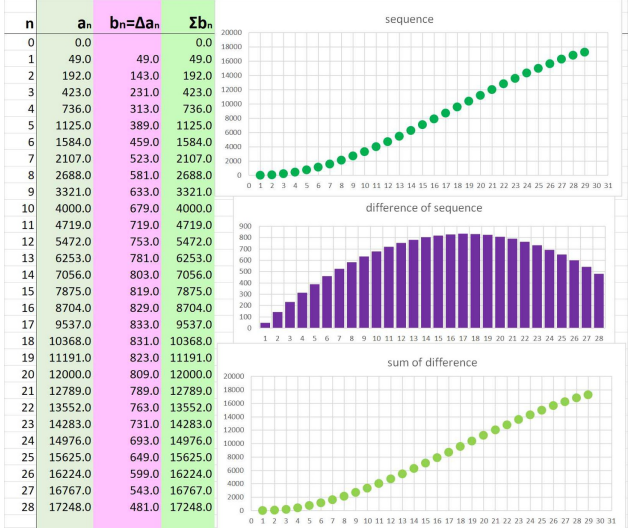
- From a 1-form to its sum:

=R[-1]C+RC[-1]

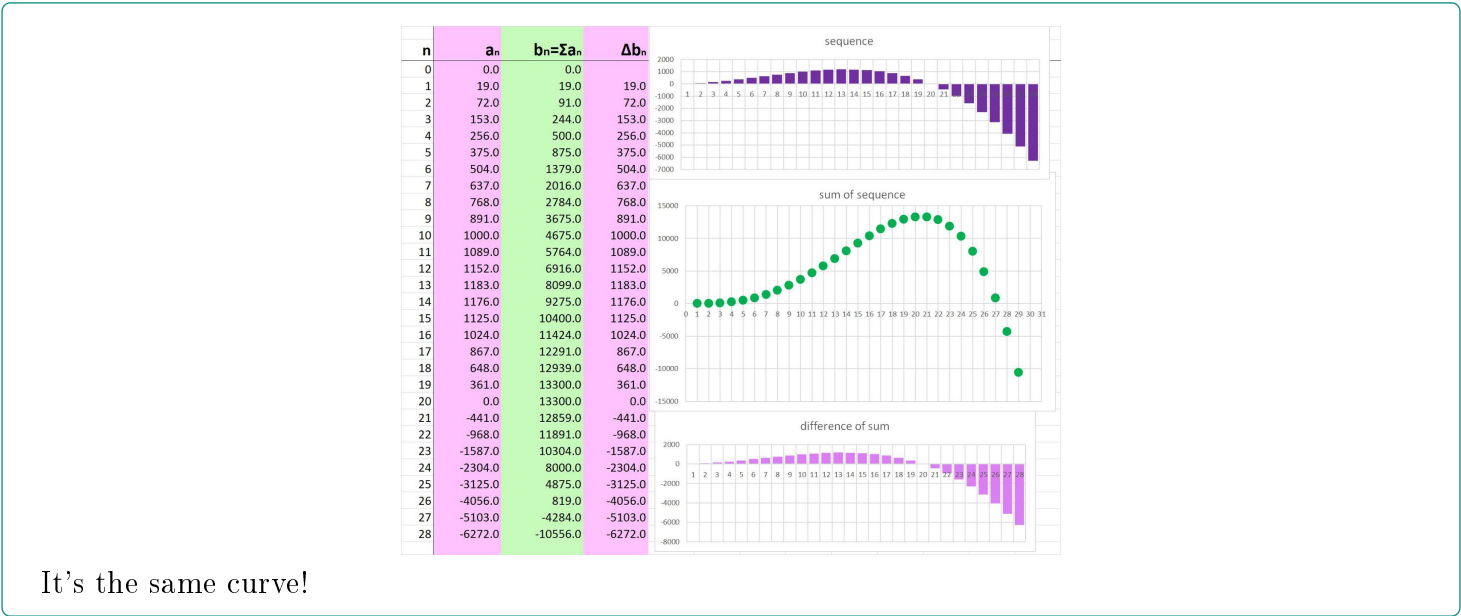
- From a 0-form to its difference:

=RC[-1]-R[-1]C[-1]

What if we combine the two consecutively? From a 0-form to its difference to the sum of the latter:

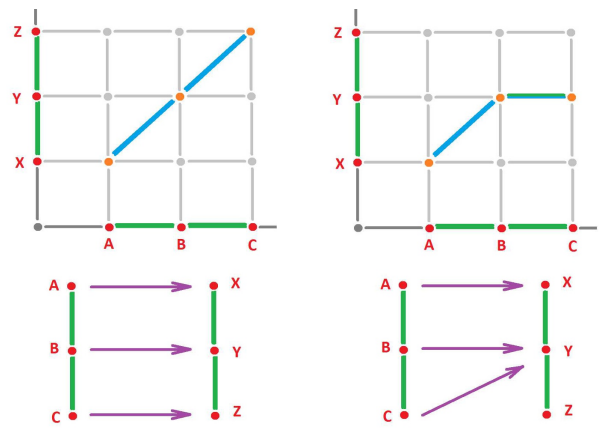


It's the same curve! Now in the opposite order, from a 1-form to its sum to the difference of the latter:



It's the same curve!

Next, this what *functions* between cell decompositions of intervals look like:



Definition 1.3.17: cell function

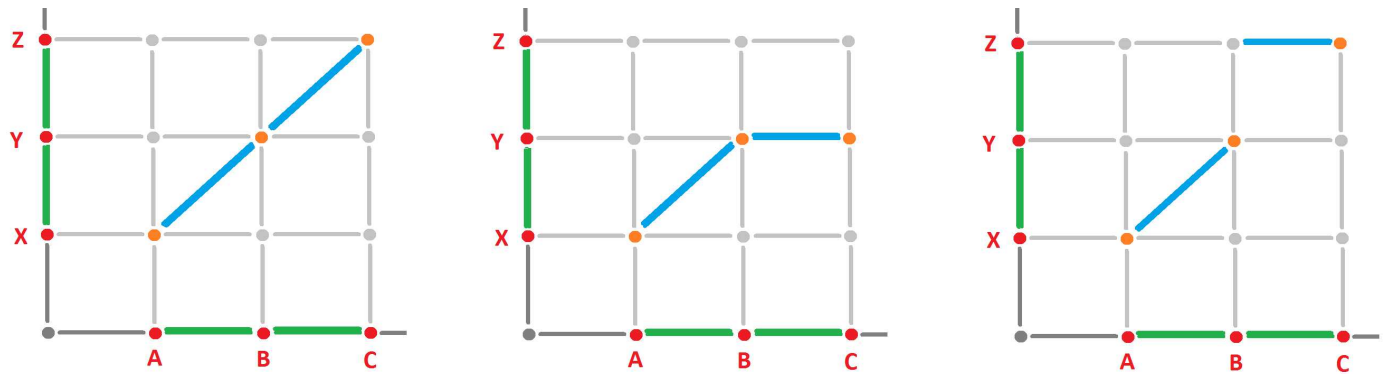
A *cell function* $y = p(x)$ is a function that assigns

- a node to each node, and
- an edge or a node to each edge,

in such a way that the end-points of each edge remain end-points:

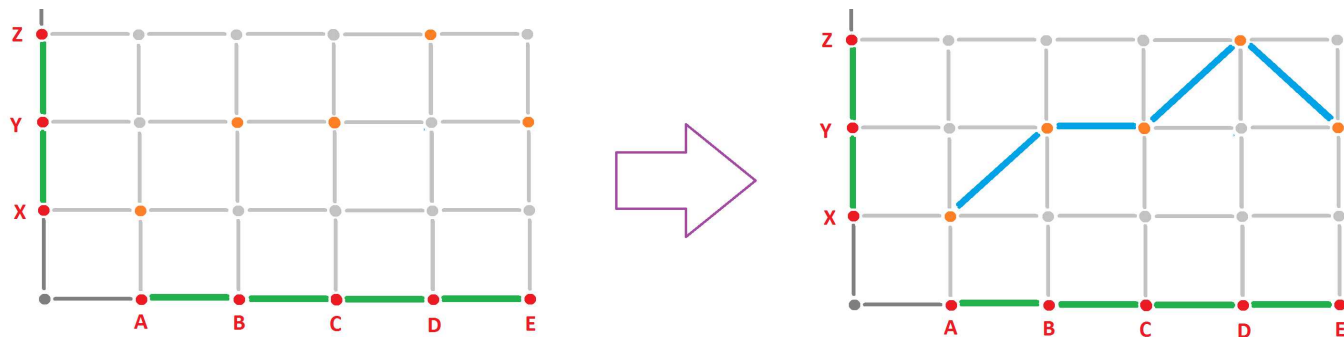
$$p([u, v]) = [p(u), p(v)]$$

The last requirement makes us avoid discontinuity:



With these, we can form compositions: $q \circ p$ of a 0- or 1-form q with another function or form p , the values of p have to be 0- and 1-cells respectively.

So, cell function p assigns a k - or $k - 1$ -cell to each k -cell:



Because of the property, the values of a cell function on the edges can be reconstructed from its values on the nodes. The former is then analogous to the *difference* of the cell function.

For convenience, we assume that Δ is zero when computed over any node x .

Theorem 1.3.18: Chain Rule

The difference of the composition of two functions is the composition of the difference of the latter with the former; i.e., for any cell function $x = p(t)$ from $[a, b]$ to $[c, d]$ and any 0-form $y = g(x)$ on $[c, d]$, we have the differences satisfy:

$$\Delta(g \circ p) = \Delta g \circ p.$$

In other words, we have for each edge s :

$$\Delta(g \circ p)(s) = \Delta g(p(s)).$$

1.4. Differential forms

They are the discrete analogs of discrete forms.

Question: Is the derivative $\frac{dy}{dx}$ a fraction?

The answer that followed the definition was an emphatic No!

A more advanced answer we give here is: Yes, here's why.

Suppose we have a function $y = f(x)$ and we are to study its behavior around a point $x = a$. The derivative at a is

$$\left. \frac{dy}{dx} \right|_{x=a} = \text{the slope of the tangent line through } (a, f(a)) = \frac{\text{rise}}{\text{run}}.$$

This *is* a fraction after all!

Example 1.4.1: quadratic

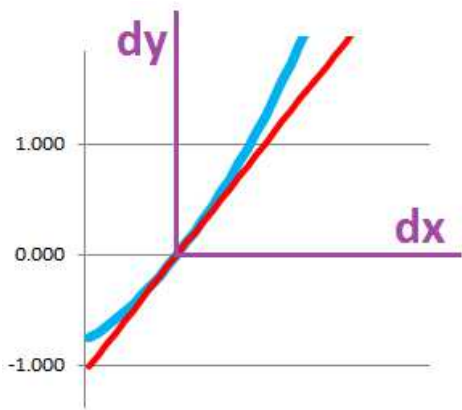
Specifically, suppose $f(x) = x^2 + 2x$. At $a = 0$, we have $f(0) = 0$, so our interest is the point $(0, 0)$. Then,

$$\left. \frac{dy}{dx} \right|_{x=0} = 2x + 2 \Big|_{x=0} = 2.$$

If this is a fraction, what would be the meaning of this:

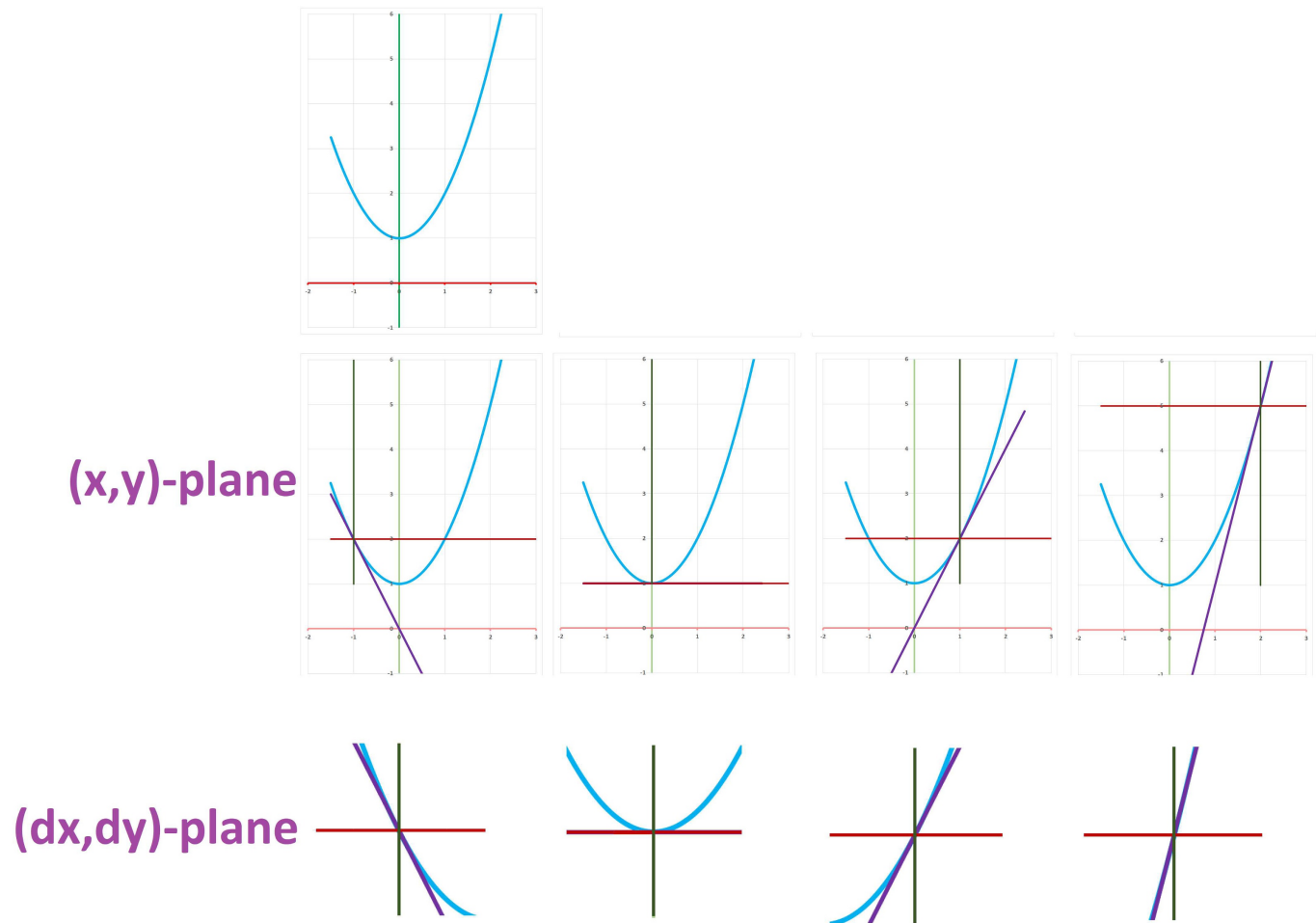
$$dy = 2 \cdot dx?$$

It is the equation of the tangent line written with respect to the two variables dy and dx :



Example 1.4.2: multiple points

Suppose we have a function $y = x^2 + 1$. We choose four points on its graph: $(-1, 2)$, $(0, 1)$, $(1, 2)$, and $(2, 5)$. At each of the points we create a “local” coordinate system (dx, dy) :



Within each of those systems the tangent line is a linear function:

point	equation
$(-1, 2)$	$dy = -2 \, dx$
$(0, 1)$	$dy = 0$
$(1, 2)$	$dy = 2 \, dx$
$(2, 5)$	$dy = 4 \, dx$

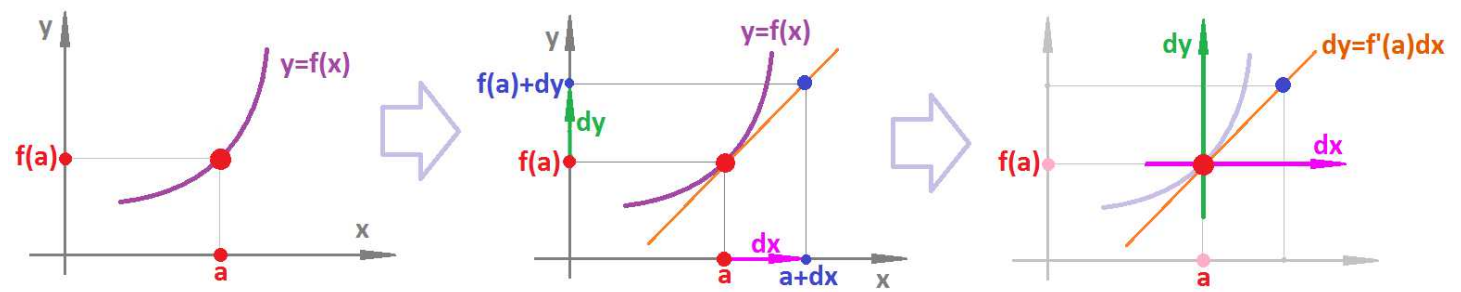
So, depending on the location, we have provided the dependence of the direction of y on the direction of x .

Thus, the equation

$$dy = f'(a) \cdot dx$$

refers to a specific location, $x = a$ and $y = f(a)$, on the xy -plane and it is a relation between the two *new variables* as the old ones have been specified.

Can we see dx , dy on the graph?



Thus, we have:

- dx is the run of the tangent line.
- dy is the rise of the tangent line.

They are called the *differentials* of x and y respectively.

Keep in mind that here dx is just a certain variable related to x (to emphasize this point, the formula can be re-written as $Y = f'(a) \cdot X$). The algebra may come from the example above:

- y depends on x via $y = f(x)$.
- dy depends on x and dx via $dy = f'(x)dx$.

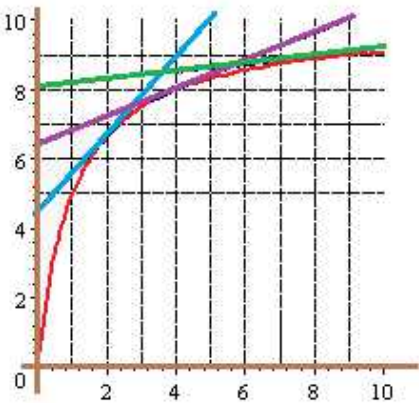
Example 1.4.3: linearization

Given a function $f(x) = x^2$, find its best linear approximation at $a = 1$. Since $f'(x) = 2x$, we see that $f'(a) = f'(1) = 2$ and, therefore, the best linear approximation of f at $a = 1$ is

$$T(x) = f(a) + f'(a)(x - a) = 1 + 2(x - a).$$

Now we interpret $x - a$ as dx . Then, if we ignore the constant part, we can write $dy = 2dx$. The equation expresses our derivative in terms of these new variables, the differentials. We capture the relation between the increment of x and that of $y - \textit{close to } a$. Indeed, y grows two times as fast as x . We acquire this information by introducing a new coordinate system (dy, dx) . In this coordinate system, the best linear approximation (given by the tangent line) becomes simply a linear function.

The analysis presented above applies to every point – and to all points at once:



Recall also from [Chapter 1PC-12](#) how we learned to look at the integral differently: we change *what* we integrate. Instead of a function,

$$\int k(x) \, dx,$$

it is a *differential form*, $k(x) \cdot dx$. As presented above, the form comes from the following:

$$y = f(x) \text{ at } x = a \implies \frac{dy}{dx} = f'(a),$$

and, furthermore,

$$\implies dy = f'(a) \cdot dx.$$

This is a relation between the two extra variables, once the relation between the old ones has been specified. The dependence between the differentials varies from location to location. So, dx is the differential of x , which is a variable separate from, but related to, x . Then, $f'(x) \cdot dx$ is just a function of two variables. The dependence of the differential form on the second variable is especially simple; it's a multiple.

Suppose we have a composition:

$$x \rightarrow u \rightarrow y.$$

Recall how the *Chain Rule*, in the Leibniz notation, is interpreted as, and it is, a “cancellation” of du (when it's not zero):

$$\frac{dy}{dx} = \frac{dy}{du} \frac{du}{dx}.$$

We also represented the formula of integration by substitution as follows:

$$\int f(u) \cdot \frac{du}{dx} \, dx = \int f(u) \, du.$$

So, under a substitution $u = g(x)$ in an integral, we also substitute:

$$du = g' \, dx.$$

Definition 1.4.4: differential 1-form

A *differential form of degree 1*, or simply a 1-form, is defined as a function of two variables:

$$\varphi = \varphi(x, dx) = g(x) \cdot dx$$

where $y = g(x)$ is a function of x , linear with respect to the second variable.

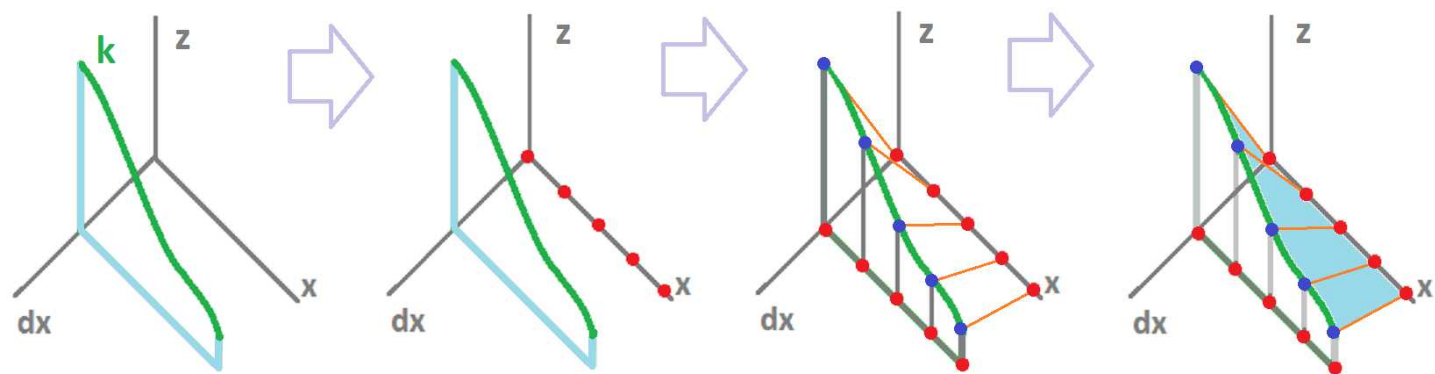
Warning!

The symbol “.” is nothing but multiplication and it is often omitted.

The point of the new concept is to make a careful distinction between the location, x , and the direction, dx . Let's plot this function:

- First we plot the graph of g (green), which is the restriction of our function φ to a fixed value of dx .
- Then we observe that φ is 0 when $dx = 0$ and plot those points on the x -axis (blue).
- Finally we connect these dots to the curve with *straight lines* (purple).

The result is this surface:



Note that the sum and the difference (but not the product or quotient) of two 1-forms is also a 1-form. Differentiation of forms is very simple.

Definition 1.4.5: differential 0-form

A *differential form of degree 0*, or simply a 0-form, is any function $y = f(x)$ of x . Its *exterior derivative* df is defined to be the 1-form given by:

$$df = f'(x) \, dx$$

The following notation is also common:

Exterior derivative

$$dy = f'(x) \, dx$$

Thus, the exterior derivative of a function contains all information about its derivative and vice versa. However, the former provides a direct answer to this question:

- If we are at $x = a$ and make a step dx , what is the step dy of y ?

Example 1.4.6: displacement

Suppose x is time and $y = f(x)$ is the location at time x . Let’s re-interpret the above question:

- If at time $x = a$ we are at $y = f(a)$ and then we move for a short time dx more, how far will we go?

The distance is velocity multiplied by time:
$$\text{Displacement} = f'(a) \cdot dx,$$
but only when the velocity, f' , is constant! In the general case, that is just an *estimate*.

Example 1.4.7: integration by substitution

We have used this algebra for *integration by substitution*. For the integral
$$\int x \sin x^2 \, dx,$$
we introduce a new variable
$$u = x^2$$
and then compute the exterior derivative of this function:
$$du = 2x \, dx.$$

Our definition of differential form treats both the integrand above and the last expression as simple cases of *multiplication* of two numbers. That is why we are at liberty to algebraically manipulate these expressions the way we have.

Integration of forms is understood in the same sense as before.

Definition 1.4.8: integral of 1-form

The *integral* of a 1-form $\varphi = g \, dx$ over an interval $[a, b]$ is defined to be

$$\int_a^b \varphi = \int_{[a,b]} \varphi = \int_a^b g(x) \, dx$$

Then the form $g \, dx$ is *integrable* whenever g is integrable.

The main application of differential forms is via their discrete counterpart discussed in the next section. We will also see applications of differential forms in the multidimensional case in the following chapters.

The following is a simple re-statement with our new notation of a familiar theorem.

Theorem 1.4.9: Fundamental Theorem of Calculus

Suppose φ is a 1-form integrable on interval $[a, b]$. Then,

$$\int_a^b \varphi = \int_{[a,b]} \varphi = F(b) - F(a)$$

for any 0-form F that satisfies:

$$dF = \varphi.$$

In order to study a real-valued function $y = f(x)$ defined on an interval, we now keep track of *two variables*:

- the locations, x vs. y , and
- the directions, dx vs. dy ,

as follows:

$$(x, dx) \mapsto (y, dy) = (f(x), f'(x)dx)$$

A function can be *sampled* in order to convert it to a discrete form. It is sampled at

- the nodes, producing a discrete 0-form, or at
- the edges, producing a discrete 1-form.

The latter, however, have an alternative, and preferred, interpretation: integration.

Theorem 1.4.10: sampling

- A differential 0-form sampled at the 0-cells is a discrete 0-form; i.e., if f is a differential 0-form, the corresponding discrete 0-form is defined by:

$$g(x_i) = f(x_i)$$

- A differential 1-form sampled at the 1-cells via integration is a discrete

1-form; i.e., if φ is a differential 1-form, the corresponding discrete 1-form is defined by:

$$s\left([x_i, x_{i+1}]\right)=\int_{\left[x_i, x_{i+1}\right]} \varphi$$

Example 1.4.11: motion

To follow the idea from the last section, the exterior derivative provides a direct answer to this question:

► Suppose x is time and $y = f(x)$ is the location at time x . If at time $x = a$ we are at $y = f(a)$ and then we move for a short time dx more, how far will we go?

The distance is velocity multiplied by time:

$$\text{Displacement} = f'(a) \cdot dx,$$

and this time the velocity, f' , is assumed to be constant throughout the interval.

1.5. Solution sets of ODEs

We will keep in mind the following idea from precalculus:

- The solution of an equation is a *set*!

For example,

$$x^2 = 1 \implies \text{solution set} = \{-1, 1\}.$$

Every differential equation will have infinitely many solutions but one solution set.

Now, suppose the velocity comes from an explicit formula as a function of the location $z = f(y)$ defined on an interval J , is there an explicit formula for the location as a function of time $y = y(t)$ defined on an interval I ? The last equation is satisfied for *every* $\Delta t > 0$ small enough. Taking the limit over $\Delta t \rightarrow 0$ gives us the following:

$$y'(t) = f(y(t)),$$

provided $y = y(t)$ is differentiable at t and $z = f(y)$ is continuous at $y(t)$. Such an equation may be *possible to solve*.

Let’s review the differential equations we have seen:

equation		a solution
1.	$f'(x) = x^2$	$\longrightarrow f(x) = x^3/3$
2.	$f'(x) = f(x)$	$\longrightarrow f(x) = e^x$

When x represents time and $y = f(x)$ represents location, the equations have simple interpretations:

1. The velocity is known at every moment of time: missile.
2. The velocity is known at every location: liquid flow.

This interpretation justifies using t for the name of the independent variable.

In the field of differential equations, it is also common *not to name the functions* anymore but use instead the names of the variables under this more compact notation:

ODEs

1. $y' = t^2 \longrightarrow y = t^3/3$ is a solution.

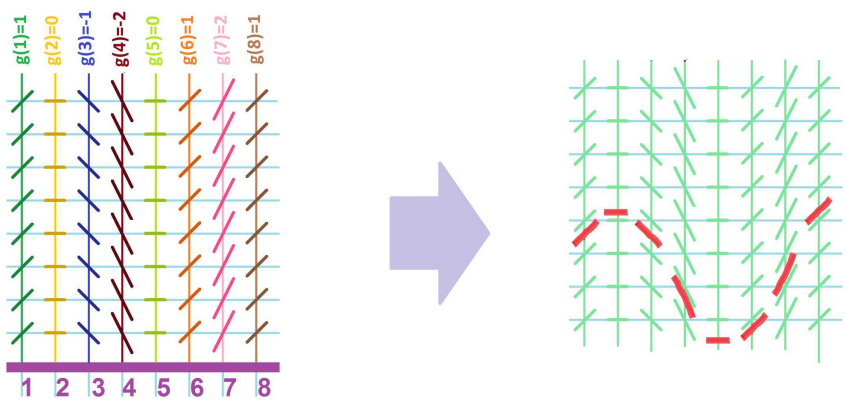
2. $y' = y \longrightarrow y = e^t$ is a solution.

They are generalized as follows.

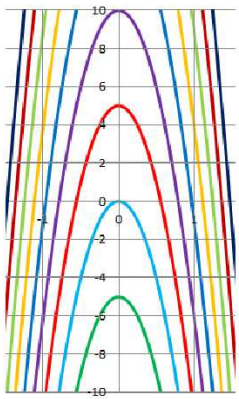
First, we have g independent of y . We are after functions $y = y(t)$ that satisfy the equation:

$$y'(t) = g(t) ,$$

for some function g . It is *location-independent*. Here, for every t , we know the slope of the tangent line at $(t, y(t))$. This slope is the same along any given *vertical* line:



When the function g happens to be constant, say -1 , its solution set is the family of these vertically shifted parabolas:



They are represented by:

$$y = -\frac{1}{2}t^2 + C ,$$

where C is any real number. Then, we have an explicit representation of the solution set:

$$\left\{ y = -\frac{1}{2}t^2 + C : C \in \mathbf{R} \right\} .$$

An examination of the picture reveals that this family of curves satisfy these two properties:

1. The plane is fully covered by the curves.
2. The curves don't intersect.

Example 1.5.1: quadratic case

Let's prove these statements.

First, “The plane is fully covered by the curves” is translated as “The plane is the union the curves” and, furthermore, “Every point on the plane belongs to one of the curves”. Suppose (t_0, y_0) is one such point. Let’s find a curve for it. This means that we need to find a specific C , that’s all. The point (t_0, y_0) belongs to the curve $y = -\frac{1}{2}t^2 + C$ means that the equation is satisfied for it:

$$y_0 = -\frac{1}{2}t_0^2 + C .$$

Solve for C :

$$C = y_0 + \frac{1}{2}t_0^2 .$$

Done!

Second, “The curves don’t intersect” is translated as “Two curves that intersect are the same curve”. Suppose we have two such curves:

$$y = -\frac{1}{2}t^2 + C ,$$

and

$$y = -\frac{1}{2}t^2 + K ,$$

for some real numbers C and K . Suppose they intersect. This means that there is a point (t_0, y_0) that belongs to both. Therefore, both of the equation are satisfied for this point:

$$y_0 = -\frac{1}{2}t_0^2 + C ,$$

and

$$y_0 = -\frac{1}{2}t_0^2 + K .$$

We solve this *system of equations* to conclude:

$$C = K .$$

That’s the same curve!

Second, we have g independent of t . We are after functions $y = y(x)$ that satisfy the equation:

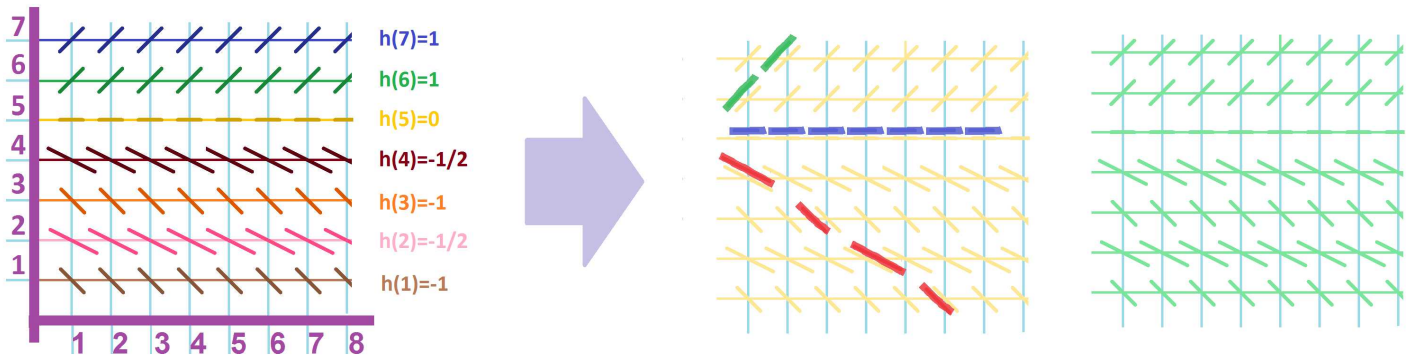
$$y'(t) = h(y(t)) ,$$

for some h .

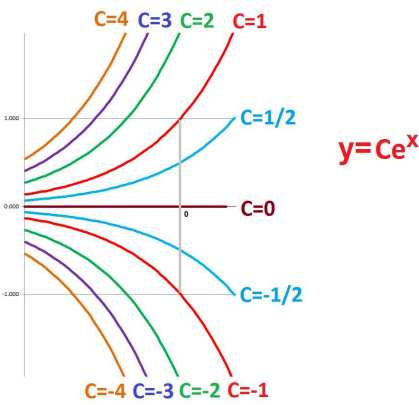
Warning!

You can’t integrate $y' = y$.

It is *time-independent*. Here, for every y , we know the slope of the tangent line at (t, y) . This slope is the same along any given *horizontal* line:



When the function $h(x) = x$ happens to be the identity, its solution is the family of vertically stretched curves:



They are represented by:

$$y = Ce^t,$$

where C is any real number. Then, we have an explicit representation of the solution set:

$$\{y = Ce^t : C \in \mathbf{R}\}.$$

An examination of the picture reveals that this family of curves satisfy these two properties:

- 1. The plane is fully covered by the curves.
- 2. The curves don't intersect.

Example 1.5.2: exponential case

Let's prove these statements.

First, "The plane is fully covered by the curves" is understood as "Every point on the plane belongs to one of the curves". Suppose (t_0, y_0) is one such point. Let's find a curve for it. This means that we need to find a specific C . The point (t_0, y_0) belongs to the curve $y = -\frac{1}{2}t^2 + C$ means that the equation is satisfied for it:

$$y_0 = Ce^{t_0}.$$

Solve for C :

$$C = y_0/e^{t_0}.$$

Done!

Second, "The curves don't intersect" is understood as "Two curves that intersect are the same curve". Suppose we have two such curves:

$$y = Ce^t,$$

and

$$y = Ke^t,$$

for some real numbers C and K . Suppose they intersect. This means that there is a point (t_0, y_0) that belongs to both. Therefore, both of the equation are satisfied for this point:

$$y_0 = Ce^{t_0},$$

and

$$y_0 = Ke^{t_0}.$$

We solve this *system of equations* to conclude:

$$C = K.$$

That's the same curve!

We considered difference equation (discrete ODE) of first order defined to be the following:

$$\frac{\Delta y}{\Delta t} = f(t, y) .$$

We take the limit of the right-hand side.

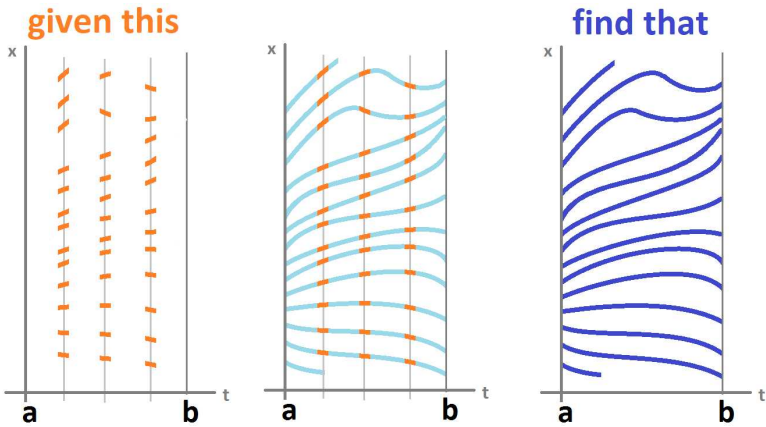
Definition 1.5.3: ODEs

Suppose f is some function of two variables. An *ordinary differential equation of first order*, or simply an ODE, is defined to be the following:

$$\frac{dy}{dt} = f(t, y) \quad \text{or} \quad y' = f(t, y)$$

Here, “ordinary” refers to the fact that there is only one independent variable and only one derivative. The “order” refers to the order of the derivative. We will refer to f and h as the “right-hand side”.

Generally, the solution set of an ODE has neither of the above patterns:



Consider the two examples of ODEs from before:

$$\begin{aligned} y' &= -gt + v_0 &\implies y(t) &= -gt^2/2 + v_0t + C \\ y' &= y &\implies y(t) &= Ce^t \end{aligned}$$

There is one solution for each C , as we know. Each of them has domain $(-\infty, \infty)$.

Unlike these two ODEs, some others have solutions that cannot be defined to the whole real line.

Example 1.5.4: asymptotes

Some of them simply run away. For example, here is a simple example of such an ODE:

$$y' = 1/x .$$

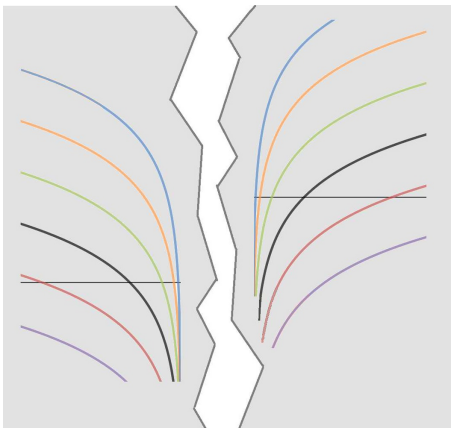
The right-hand side function,

$$f(x, y) = 1/x ,$$

is undefined at $x = 0$. For a function of two variables, its domain consists now of two open half-planes:

$$x < 0 \quad \text{and} \quad x > 0 .$$

Therefore no solution to our ODE can cross the line $x = 0$.



We have, in fact, two families of solutions:

$$\begin{aligned} y &= \ln(-x) + C && \text{for } x > 0; \\ y &= \ln(x) + K && \text{for } x < 0. \end{aligned}$$

They exist separately from each other and should not be combined arbitrarily.

This is the reason why we take the following point of view on solutions.

Definition 1.5.5: solution of ODE

A *solution* of ODE $y' = f(t, y)$ is a function $y = y(t)$ differentiable on an open interval I such that for every t in I we have:

$$y'(t) = f(t, y(t))$$

It may also be called a *strong solution*.

Note that we require the domain to be an *open* interval so that the differentiability is properly defined.

Let's integrate the equation over some interval $[a, x]$ with variable right end:

$$\int_a^x y'(t) \, dt = \int_a^x f(t, y(t)) \, dt.$$

According to the Fundamental Theorem of Calculus, the left-hand side simplifies:

$$y(x) - y(a) = \int_a^x f(t, y(t)) \, dt$$

We have an *integral equation*.

Definition 1.5.6: weak solution of ODE

A *weak solution* of ODE $y' = f(t, y)$ is a function y continuous on a closed interval $[a, b]$ such that for every x in $(a, b]$, the integral equation is satisfied.

Note that we require the domain to be an *closed* interval so that we can integrate.

Example 1.5.7: step-function

In the later case the function can be a step-function. For example, this is how easy it is to produce the solutions:

function f

-1	-1	0	0
0	1	0	-1
1	1	0	0
0	1	-1	-1

slopes

solutions y

Such functions are called “piece-wise linear”.

Exercise 1.5.8

Do any of the solutions intersect?

Exercise 1.5.9

Plot a few solutions for the following right-hand side function:

1	1	-1	0
0	1	-1	-1
2	1	-2	0
-1	1	0	2

Above, we showed the following.

Theorem 1.5.10: weak solutions

Every (strong) solution is a weak solution.

Example 1.5.11: converse

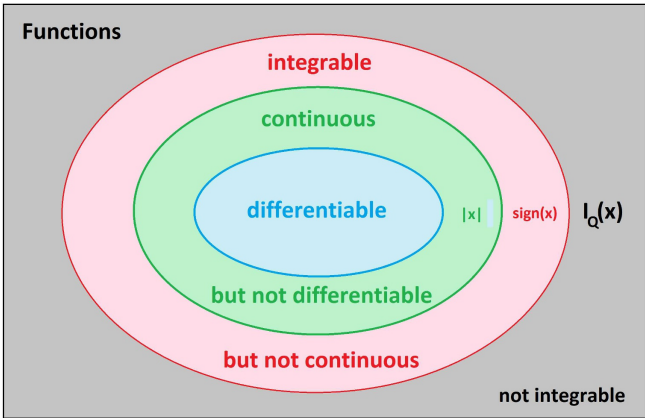
The converse is false. Let’s choose the right-hand function f to be the sign function:

Then an “obvious” solution is the absolute value:

The differentiation fails at $t = 0$ however:

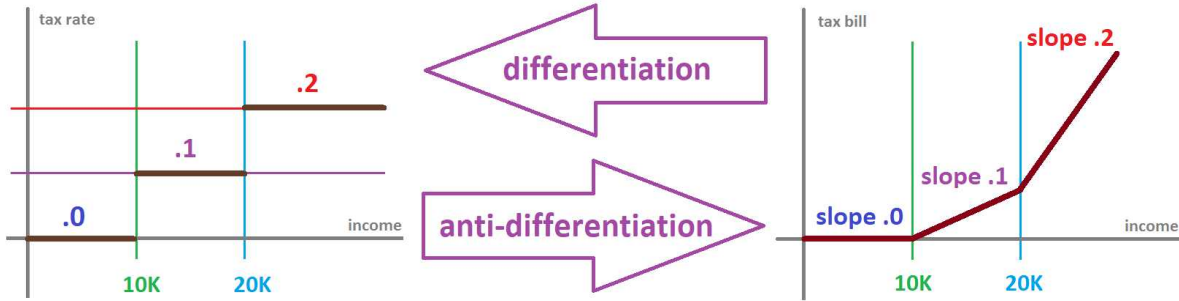
$$y' = \text{sign}(t), \quad y = |t|.$$

Continuous vs. differentiable vs. integrable:

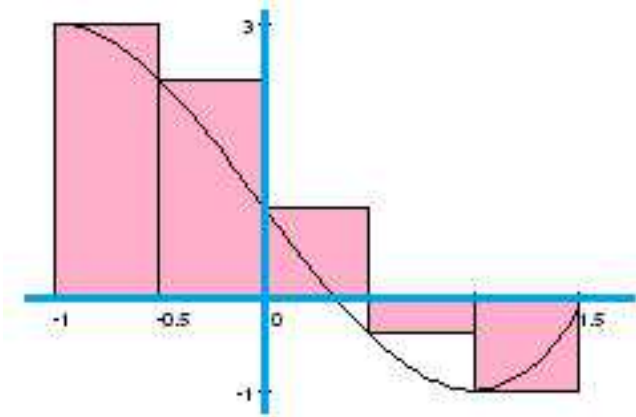


Example 1.5.12: incremental change

Weak solution may emerge from processes guided by functions that change incrementally. One example is how the income tax emerges from the income tax rate:

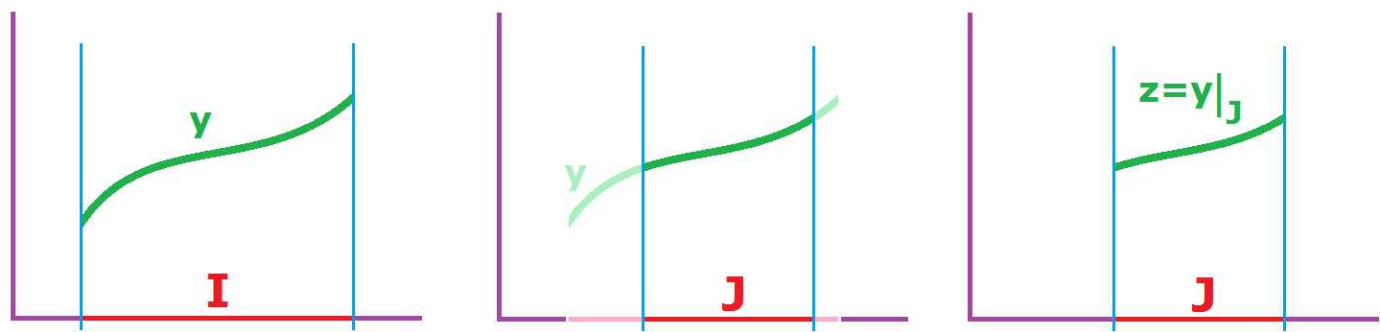


We have the ODE on the left and a solution on the right. It is solved by a simple integration. We ignore the non-differentiability at the ends of the brackets, but the continuity of the weak solution is still required! Another source is signal processing. Weak solutions may also come from sampling of the right hand side function of an ODE as a way to solve it numerically:



The function is replaced with a step-function.

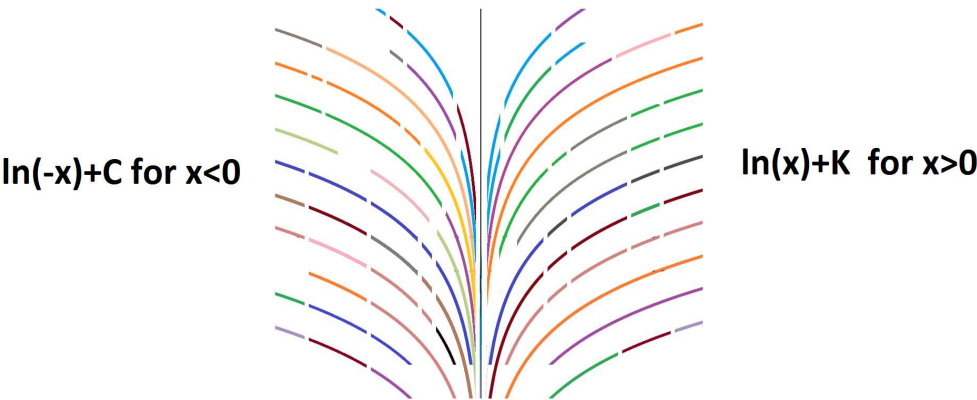
The very first observation we should make is that, according to the definition, there is a separate solution for each interval inside I , i.e., if y is a solution on interval I then so is its restriction $z = y|_J$ to any (open or closed respectively) interval J in I .



This new function is defined very simply:

$$z(t) = y(t) \text{ for each } t \text{ in } J.$$

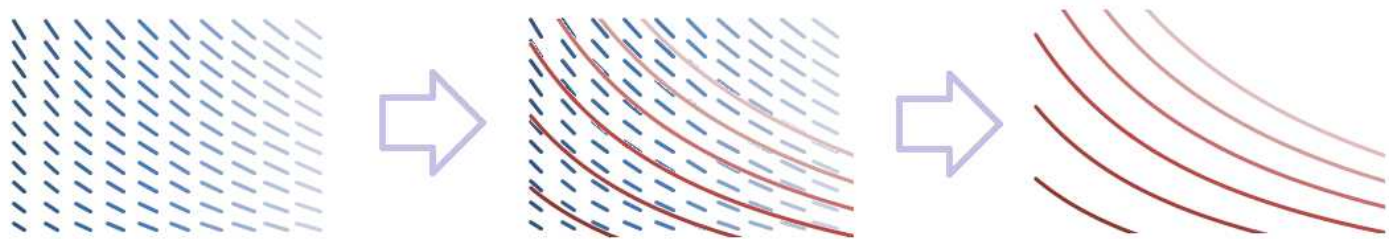
For the above example, each segment of a curve seen below is a solution:



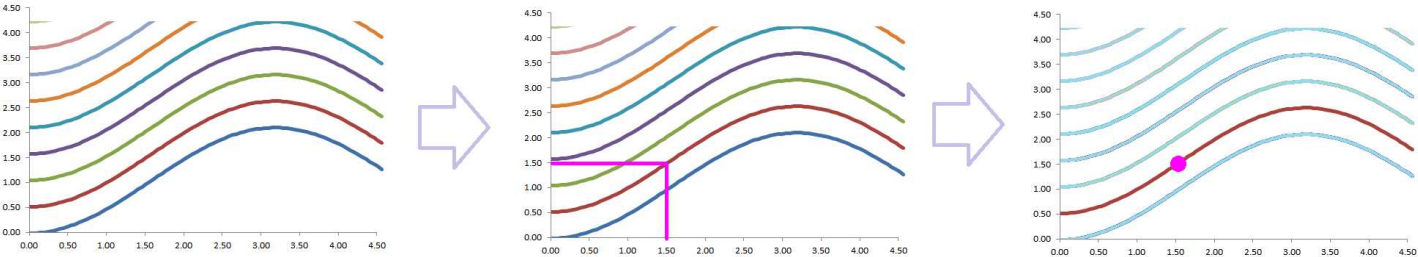
Warning!

The interval I may be infinite:
$$I = (-\infty, +\infty), (-\infty, b), (a, +\infty).$$

Recall that solving an ODE means to face a “field of slopes” and then search for curves with these tangents:



Solving problems about the fall of a ball thrown from a particular location at a particular time required picking out one of them:



There is a related concept in the theory of ODEs.

Definition 1.5.13: initial value problem

For a given ODE $y' = f(t, y)$ and a given pair of values (t_0, y_0) , the *initial value problem*, or an IVP, is

$$y' = f(t, y), \quad y(t_0) = y_0$$

and its *solution* is a solution of the ODE that satisfies the *initial condition*, $y(t_0) = y_0$.

Warning!

A different initial value condition may produce the same solution.

In the example in the last section, the initial value problem is solvable for each (t_0, y_0) in the domain. That is why the curves fill the plane. In the last example, they fill all plane except the line $t = 0$.

The first property of the solution set is:

- The curves cover the plane.

Definition 1.5.14: existence

We say that an ODE satisfies the *existence* property at a point (t_0, y_0) when the IVP:

$$y' = f(t, y), \quad y(t_0) = y_0,$$

has a solution.

In other words, for every initial condition there *exists* at least one solution that starts there.

It doesn't matter how small is the domain of this solution.

There are no (strong or weak) solutions passing through the vertical line:



If your model of a real-life process doesn't satisfy this property, it may be inadequate. It is as if the process starts but never continues.

From what we know about antiderivatives, we conclude the following.

Theorem 1.5.15: Existence for Location-Independent ODE

An ODE the right-hand side of which is a function independent of y and integrable with respect to t on an open interval I :

$$y' = f(t),$$

satisfies the existence property for every point (t_0, y_0) with t_0 in I .

Proof.

Indeed, the solution is just the appropriate antiderivative of f .

The second property of the solution set is:

- The curves don't intersect.

Definition 1.5.16: uniqueness

We say that an ODE satisfies the *uniqueness* property at a point (t_0, y_0) if every pair of solutions, y_1, y_2 , of the IVP:

$$y' = f(t, y), \quad y(t_0) = y_0$$

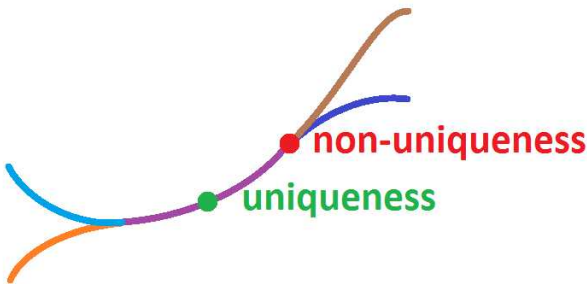
are equal,

$$y_1(t) = y_2(t)$$

for every t in some open interval that contains t_0 .

In other words, for every initial condition there is at most one solution, i.e., *unique* solution, that starts there.

To violate uniqueness, the two strong solutions will have to have the same slope (i.e., they are tangent to each other) at the fork:



Example 1.5.17: non-uniqueness

Weak solutions can easily part ways:

1

-1

→

→

The definition says that their restrictions to some interval J are equal:

$$y_1 \Big|_J = y_2 \Big|_J .$$

If your model of a real-life process doesn't satisfy this property, it may be inadequate. It's as if you have all the data but can't predict even the nearest future.

From what we know about antiderivatives, we conclude the following.

Theorem 1.5.18: Uniqueness for Location-Independent ODE

An ODE the right-hand side of which is a function independent of y and integrable with respect to t on an open interval I :

$$y' = f(t),$$

satisfies the uniqueness property for every point (t_0, y_0) with t_0 in I .

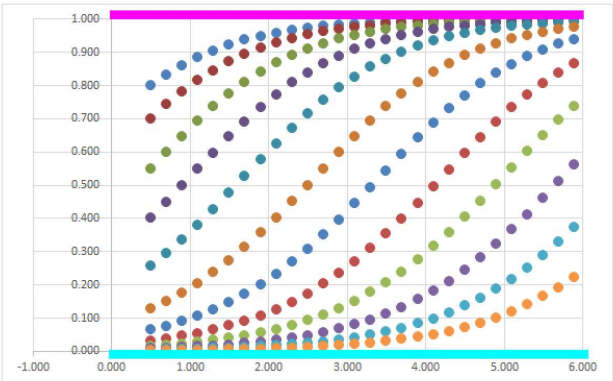
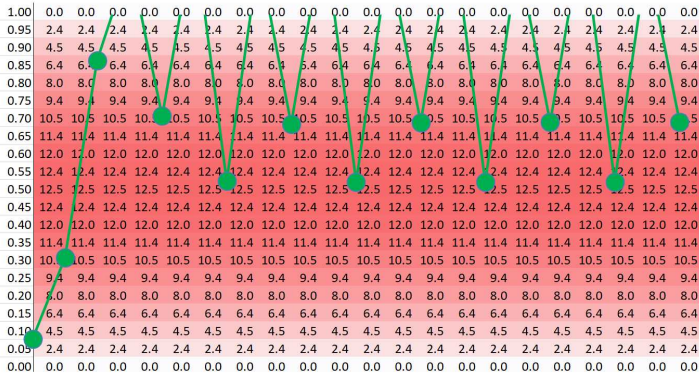
Proof.

Indeed, the solution is always one of the antiderivatives of f . We also know that these antiderivatives always differ by a constant and, therefore, their graphs never intersect.

We will see ODEs without the uniqueness property later.

Example 1.5.19: logistic equation

For the logistic equation, the difference equation $\Delta y = ky(1 - y)\Delta t$ may produce strange results (left) but not the corresponding ODE $y' = ky(1 - y)$ (right):



One can show that this is true by appealing to the uniqueness of the solutions: they can't cross the trivial solutions $y = 0$ and $y = 1$.

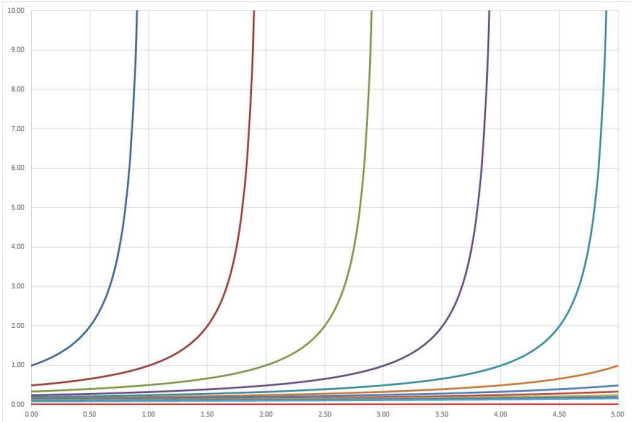
So, for a model to be valid, it has to satisfy:

- 1. Existence: The plane is fully covered by the solution curves.
- 2. Uniqueness: The solution curves don't intersect.

There is another important condition!

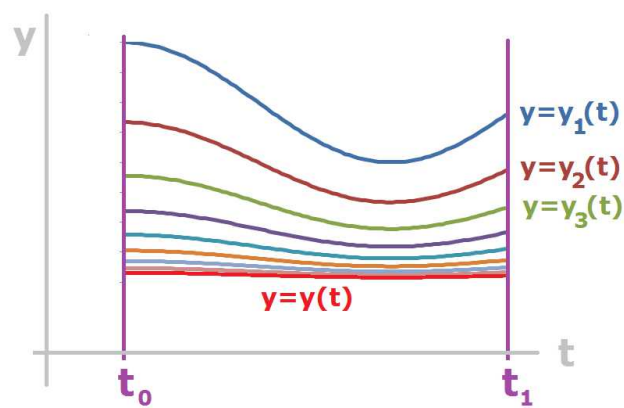
Example 1.5.20: weather forecast

It is well known that a weather forecast beyond 3 days is very unreliable. A situation like that occurs when small change in the input produces a large change in the output:



Suppose the forecast depends on the measuring the current temperature. In an attempt to improve the accuracy of the forecast, we make this measurement more and more precise. However, the results will eventually run away from the nominal, $y = 0$.

What we are after is models that produce solutions that approach the nominal as the initial condition approach the nominal:



This is what we may require in order to avoid this kind of instability.

Definition 1.5.21: continuity of IVP

We say that an ODE satisfies the *continuity of IVP* property at (t_0, y_0) if for each solution y of the IVP, any $t_1 > t_0$ within the domain of y , and any sequence of solutions, y_n defined on $[t_0, t_1]$, of the ODE we have, as $n \rightarrow \infty$,

$$y_n(t_0) \rightarrow y(t_0) \implies y_n(t_1) \rightarrow y(t_1).$$

Such an ODE can also be called *stable*.

Example 1.5.22: parabolas

Let's take this family of parabolas again:

$$y = -t^2/2 + C.$$

A sequence of solutions will look like this:

$$y_n(t) = -t^2/2 + C_n, \quad n = 1, 2, 3...$$

Then, in particular, we have at one end, $t = t_0$:

$$y_n(t_0) = -t_0^2/2 + C_n, \quad n = 1, 2, 3...$$

If this sequence of numbers converges,

$$y_n(t_0) = -t_0^2/2 + C_n \rightarrow y(t_0) = -t_0^2/2 + C,$$

then $C_n \rightarrow C$ and, therefore, we have convergence on the other end, $t = t_1$:

$$y_n(t_1) = -t_1^2/2 + C_n \rightarrow y(t_1) = -t_1^2/2 + C.$$

Exercise 1.5.23

Prove the property for the family of exponential functions $y = Ce^t$.

Exercise 1.5.24

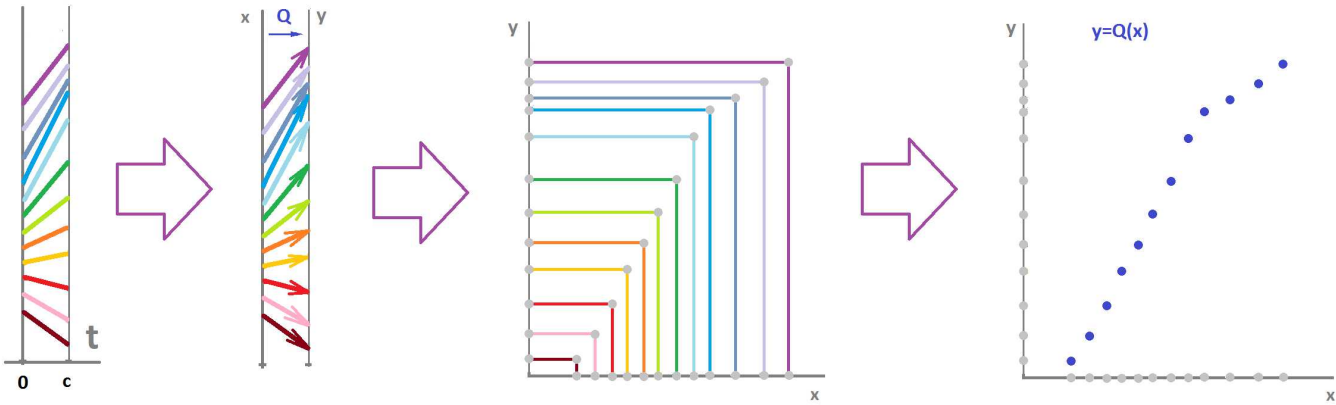
What about convergence of functions $y_n \rightarrow y$? What about the uniform convergence?

Example 1.5.25: continuity?

In order to explain the term “continuity”, let’s define a function Q of the y -axis to itself by:

$$Q(z) = y(t_1) \text{ when } y(t_0) = z.$$

It’s as if the flow rearranges the molecules. We then build the graph of this function:



It is continuous.

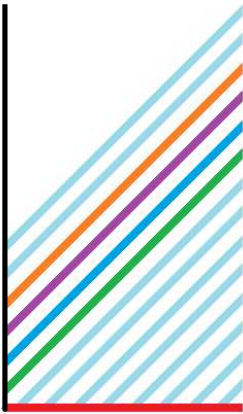
Exercise 1.5.26

Prove that the function is continuous.

If your model of a real-life process doesn’t satisfy this property, it may be inadequate. It’s as if any error in the initial data – no matter how small – may cause a significant error in your prediction.

Example 1.5.27: discontinuity of IVP

Weak solution can easily violate this property:



From what we know about antiderivatives, we conclude the following.

Theorem 1.5.28: Continuity of IVP for Location-Independent ODE

An ODE the right-hand side of which is a function independent of y and integrable with respect to t on an open interval I :

$$y' = f(t),$$

satisfies the continuity of IVP property for every point (t_0, y_0) with t_0 in I .

Proof.

Indeed, the solution is always one of the antiderivatives of f . We also know that these antiderivatives always differ by a constant and, therefore, converge to each other when this constant goes to zero.

Next, ODEs produce families of curves as the sets of their solutions... and vice versa: If a family of curves is given by an equation with a single parameter C , then differentiating the equation in a particular way will produce an ODE.

Example 1.5.29: vertically shifted

First, the family of vertically shifted graphs,

$$y = g(t) + C,$$

creates, when differentiated, an ODE:

$$y' = g'(t).$$

Second, we consider the family of vertically stretched exponential graphs,

$$y = Ce^t.$$

When differentiated, the resulting equation still has C however! To creates an ODE without C , let's first solve for C and then differentiate:

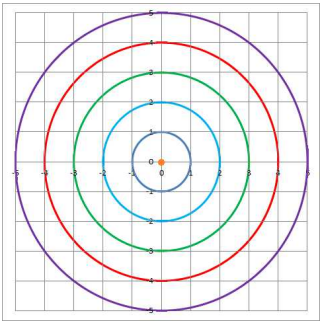
$$C = ye^{-t} \implies 0 = y'e^{-t} + y(e^{-t})' = y'e^{-t} - ye^{-t} = (y' - y)e^{-t}.$$

Since $e^{-t} > 0$, we have:

$$y' - y = 0.$$

Example 1.5.30: circles

What about these concentric circles?



They are given by

$$x^2 + y^2 = C > 0.$$

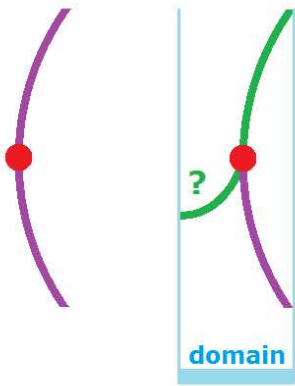
We differentiate (implicitly) with respect to x :

$$2x + 2yy' = 0,$$

or,

$$y' = -\frac{x}{y}.$$

The domain of the right-hand side is $y \neq 0$. Therefore, the existence property cannot be satisfied on the x -axis. Furthermore, even if this function was defined on the x -axis, the existence property would still break down:



As you can see, we cannot extend the solution to the left of this point. Similarly, we cannot extend the solution to the right of the point if it is on the other side of 0.

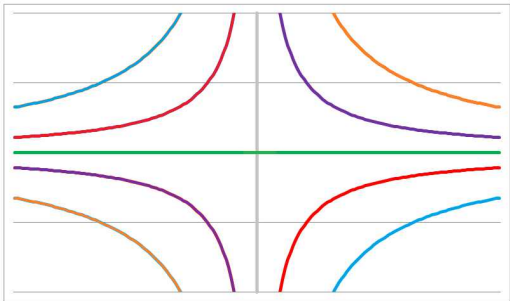
Exercise 1.5.31

What about uniqueness?

Example 1.5.32: hyperbolas

These hyperbolas are given by these equations:

$$xy = C .$$



(In the case when $C = 0$, we have the two axes.) Then, we have an ODE:

$$y + xy' = 0 ,$$

or,

$$y' = \frac{y}{x} .$$

The domain of the right-hand side is $x \neq 0$. Therefore, the existence property cannot be satisfied on the y -axis. Furthermore, even if this function was defined on the y -axis, the existence property would still break down:



We cannot have a solution passing through the y -axis.

It appears that the presence or the absence of the existence property depends on the continuity condition of the right-hand side function of the ODE. The proof of the following important theorem lies outside the scope of this book.

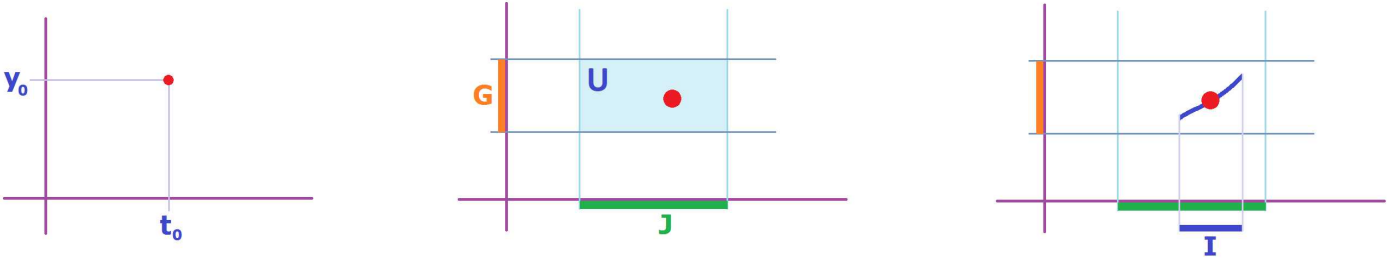
Theorem 1.5.33: Existence

Suppose U is an open set on the plane (t, y) that contains (t_0, y_0) and suppose that a function $z = f(t, y)$ of two variables defined on U is

- continuous with respect to t , and
- continuous with respect to y .

Then the ODE $y' = f(t, y)$ satisfies the existence property at (t_0, y_0) .

At its simplest, this open set U is a rectangle, the product of two intervals J and G :



Warning!

The domain of the solution doesn't have to be the whole J because it may exit the rectangle through its top or its bottom:

Exercise 1.5.34

What if $G = (-\infty, \infty)$?

Example 1.5.35: non-uniqueness

Consider this ODE:

$$\frac{dy}{dx} = y^{2/3}.$$

Notice that there is one more solution passing through $(0,0)$ in addition to the trivial solution $y = 0$:

It is

$$y = \begin{cases} 0 & \text{for } x < 0, \\ \frac{1}{27}x^3 & \text{for } x \geq 0. \end{cases}$$

What makes the right-hand side $y^{2/3}$ special in comparison to the previous examples? Its derivative is infinite at $y = 0$.

It appears that the presence or the absence of the uniqueness property depends on the differentiability condition of the right-hand side function of the ODE. The proof of the following important theorem lies outside the scope of this book. The set-up is identical to that of the last theorem except:

- continuity for y is replaced with differentiability.

Theorem 1.5.36: Uniqueness (+Existence)

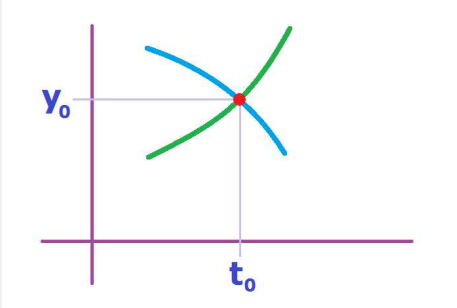
Suppose U is an open set on the plane (t, y) that contains (t_0, y_0) and suppose that a function $z = f(t, y)$ of two variables defined on U is

- continuous with respect to t , and
- differentiable with respect to y .

Then the ODE $y' = f(t, y)$ satisfies the uniqueness property at (t_0, y_0) .

Warning!

This is *not* what non-uniqueness looks like:



It appears that the presence or the absence of the continuity of IVP property depends on the differentiability condition of the right-hand side function of the ODE. The proof of the following important theorem lies outside the scope of this book. The set-up is identical to that of the last theorem.

Theorem 1.5.37: Continuity of IVP (+Uniqueness+Existence)

Suppose U is an open set on the plane (t, y) that contains (t_0, y_0) and suppose that a function $z = f(t, y)$ of two variables defined on U is

- continuous with respect to t , and
- differentiable with respect to y .

Then the ODE $y' = f(t, y)$ satisfies the continuity of IVP property at (t_0, y_0) .

Next, once we have a solution $y = y(t)$ on interval I , any restriction $z = y|_J$ of function y to any interval J inside I is also a solution. Conversely, once we have a solution z on interval J , there may be an *extension* y of z , i.e., $y = y(t)$ is a solution and $z = y|_J$. Sometimes it is impossible to further extend our solution.

Definition 1.5.38: maximal solution

A solution $y = y(t)$ is called a *maximal solution* if it doesn't have extensions.

As we have seen above, a maximal solution is often defined on $(-\infty, +\infty)$ but also may be defined on other open intervals.

Below, we will refer to maximal solutions as simply *solutions*.

1.6. Separation of variables in ODEs

Some types of ODEs can be solved explicitly.

Differential forms allow us sometimes to separate x , and dx , from y , and dy .

Example 1.6.1: integration

We re-write the ODE as an equation of differential forms and then integrate both sides:

$$\frac{dy}{dx} = x^2 \implies dy = x^2 dx \implies \int dy = \int x^2 dx \implies y + C = x^3/3 + K.$$

The two indefinite integration produce two constants. Since both are arbitrary, just one is sufficient ($Q = K - C$):

$$y = x^3/3 + Q.$$

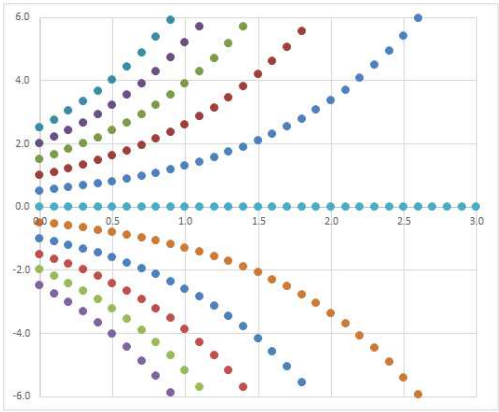
The ODE is solved.

Example 1.6.2: population model

We consider the familiar ODE:

$$\frac{dy}{dx} = y.$$

Recall that plotting its corresponding difference equation (discrete ODE) produces this:



Following the same procedure, we integrate both sides of an equation of differential forms keeping only one constant:

When $y > 0$: $\frac{dy}{dx} = y \implies \frac{dy}{y} = dx \implies \int \frac{dy}{y} = \int dx \implies \ln y = x + Q.$

This is a part of the solution set of the ODE given implicitly. It is a family of relations. Similarly, we obtain:

When $y < 0$: $\ln(-y) = x + P.$

What's left is the case of $y = 0$. That's a whole (constant) solution:

When $y = 0$: $\frac{dy}{dx} = 0 \implies y = \text{constant} \implies y = 0.$

We verify this fact by substitution.

Now explicit the solutions. We apply the logarithm to both sides of the implicit equations. First:

When $y > 0$: $\ln y = x + Q \implies y = e^{x+Q} = e^Q e^x, \quad Q \text{ real.}$

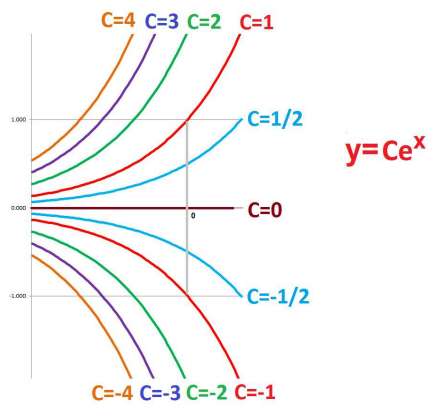
These are all positive multiples of the exponential function. Second:

When $y > 0$: $\ln(-y) = x + P \implies y = -e^{x+P} = -e^P e^x$, P real.

These are all negative multiples of the exponential function. Of course, $y = 0$ is the zero multiple. We have them all in one formula:

$y = Ce^x$, C real.

The solution set this result produces is familiar:



In general, the method applies whenever our right-hand side function splits into a factor of two functions with one depending only on x and the other only on y :

$\frac{dy}{dx} = p(x)q(y) \implies \frac{dy}{q(y)} = p(x) dx .$

Such ODEs are called *separable*.

Exercise 1.6.3

Give examples of separable and non-separable ODEs.

Theorem 1.6.4: Solutions of Separable ODEs

Every solution $y = y(x)$ of a separable ODE $y' = p(x)q(y)$ satisfies the equation:

$\int \frac{dy}{q(y)} = \int p(x) dx$

Warning!

There is no change of variables here as these are the same variables.

The question then remains whether the two integrals can be evaluated in an elementary fashion.

Example 1.6.5: algebra

Consider this ODE:

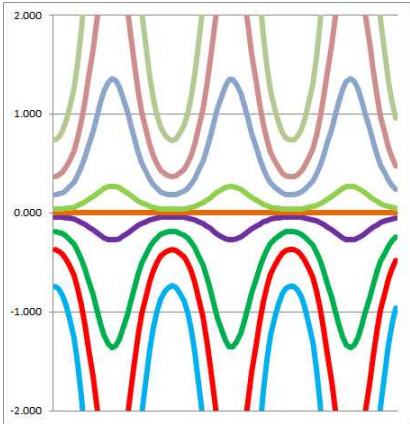
$\frac{dy}{dx} = y \sin x .$

We know that the existence and the uniqueness are satisfied.

We can also take the corresponding the difference equation:

$$\frac{\Delta y}{\Delta x} = y \sin x .$$

and plot the solutions:



Now the algebra. The ODE is separable:

$$\frac{dy}{dx} = y \sin x$$
$$\implies \frac{dy}{y} = \sin x \, dx$$
$$\implies \int \frac{dy}{y} = \int \sin x \, dx$$
$$\implies \ln y = -\cos x + Q \text{ for } y > 0, \quad \ln(-y) = -\cos x + P \text{ for } y < 0$$
$$\implies y = e^{-\cos x + Q} \text{ for } y > 0, \quad y = -e^{-\cos x + P} \text{ for } y < 0$$
$$\implies y = e^{-\cos x} e^Q \text{ for } y > 0, \quad y = -e^{-\cos x} e^P \text{ for } y < 0$$
$$\implies y = C e^{-\cos x}$$

Even though some of the solutions appear to touch the x -axis, the uniqueness is satisfied. Indeed, every solution is just a vertically stretched of the simplest one:

$$y = e^{-\cos x} .$$

Exercise 1.6.6

Prove the last statement.

Let’s apply the method to an ODE that we will see later.

Theorem 1.6.7: Linear ODE

The ODE

$$y' = a(x)y ,$$

with an integrable function a , satisfies the existence and uniqueness. The solutions of the ODE are given by:

$$y = C e^{A(x)}$$

where A is any antiderivative of a :

$$A(x) = \int a(x) \, dx$$

Thus the solution set splits into three parts:

(1) $y = Ke^{A(x)}$ with $K < 0$, (2) $y = 0$, (3) $y = Ce^{A(x)}$ with $C > 0$.

All are multiples of $y = e^{A(x)}$.

Exercise 1.6.8

Prove the theorem.

Example 1.6.9: generic

Consider this ODE:

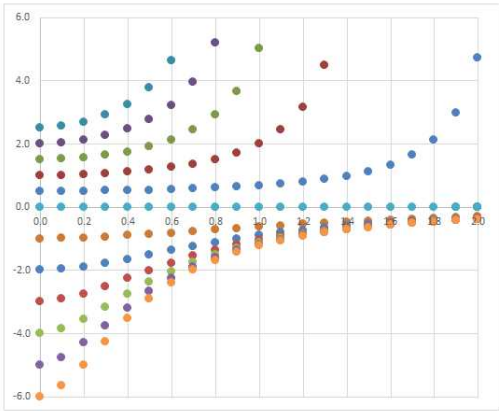
$$\frac{dy}{dx} = y^2 x .$$

We know that the existence and the uniqueness are satisfied.

We can also take the corresponding the difference equation:

$$\frac{\Delta y}{\Delta x} = y \sin x .$$

and plot the solutions:



Now the algebra. The ODE is separable:

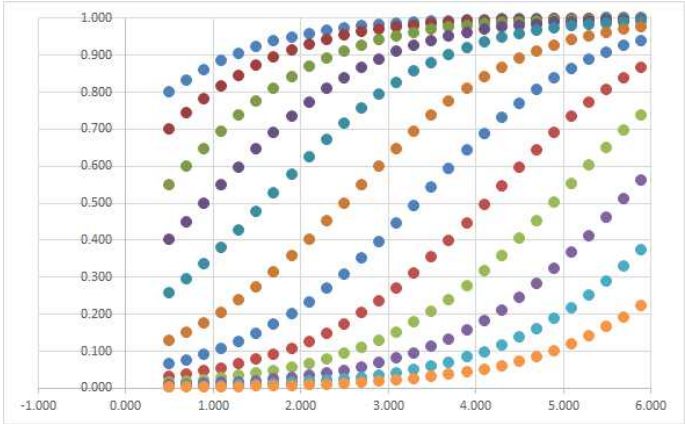
$$\begin{aligned} \frac{dy}{dx} &= y^2 x \\ \Rightarrow \frac{dy}{y^2} &= x \, dx \\ \Rightarrow \int \frac{dy}{y^2} &= \int x \, dx \\ \Rightarrow -\frac{1}{y} &= \frac{1}{2}x^2 + Q \\ \Rightarrow y &= -\frac{1}{\frac{1}{2}x^2 + Q} \text{ on each open interval of } x . \end{aligned}$$

Example 1.6.10: logistic model

We consider the familiar ODE:

$$\frac{dy}{dx} = y(1 - y) .$$

Recall that plotting its corresponding difference equation produces this:



Following the same procedure, we integrate both sides of an equation of differential forms keeping only one constant:

$$\frac{dy}{dx} = y(1 - y) \implies \frac{dy}{y(1 - y)} = dx \implies \int \frac{dy}{y(1 - y)} = \int dx \implies \frac{y}{y - 1} = x + Q \text{ for } 0 < y < 1 .$$

This is a part of the solution set of the ODE given implicitly as a family of relations.

Exercise 1.6.11

Find the explicit solutions.

1.7. The method of integrating factors

Example 1.7.1: non-separable equation

Let’s take a careful look at the left-hand side of the following (non-separable) equation:

$$y' \cdot \sin x + y \cdot \cos x = xe^{x^2} .$$

We see two pairs of functions present:

- y and its derivative y' , and
- $\sin x$ and its derivative $\cos x$.

They are “cross-multiplied”. The expression is familiar; it’s the outcome of the Product Rule:

$$(y \cdot \sin x)' = y' \cdot \sin x + y \cdot \cos x .$$

Therefore, our ODE turns into:

$$(y \cdot \sin x)' = xe^{x^2} .$$

We integrate:

$$\int (y \cdot \sin x)' \, dx = \int xe^{x^2} \, dx ,$$

and apply the Fundamental Theorem of Calculus:

$$y \cdot \sin x = \frac{1}{2}e^{x^2} + C .$$

We were lucky.

Example 1.7.2: further steps

Consider the ODE:

$$y' + y = x .$$

The left-hand side is *not* the outcome of the Product Rule... Can we make it so? After all, the pair y and y' is already there. Unfortunately, what's left doesn't work:

$$y' \cdot 1 + y \cdot 1 = x .$$

We can't just replace the 1's with some functions F and its derivative F' , can we? We can if this is the *same* function:

$$(e^x)' = e^x .$$

We multiply both sides of the equation by this factor $F(x) = e^x$:

$$y' \cdot e^x + y \cdot e^x = xe^x ,$$

or

$$y' \cdot e^x + y \cdot (e^x)' = xe^x .$$

By the Product Rule we have instead:

$$(y \cdot e^x)' = xe^x .$$

After integration, we have

$$y \cdot e^x = -e^x + xe^x + C ,$$

or even simpler:

$$y = -1 + x + Ce^{-x} .$$

Then, it is promising for this approach to have both y and y' , but not y^2 or $\sin y'$. In other words, an equation should be *linear* with respect to y and y' .

Example 1.7.3: non-homogeneous

We make a small adjustment to the ODE:

$$y' + 2y = x .$$

The left-hand side is not the outcome of the Product Rule... even if we multiply by e^x :

$$y' \cdot e^x + 2y \cdot e^x = xe^x .$$

What choice of a factor F would work:

$$y' \cdot F(x) + 2y \cdot F(x) = xF(x)?$$

We'd need this:

$$F' = 2F .$$

But that's just another ODE! And a familiar one too:

$$F(x) = e^{2x} .$$

Thus we have:

$$y' \cdot e^{2x} + 2y \cdot e^{2x} = xe^{2x} \, ,$$

or

$$y' \cdot e^{2x} + y \cdot (e^{2x})' = xe^{2x} \, .$$

Then,

$$(y \cdot e^{2x})' = xe^{2x} \, .$$

After integration, we have

$$y \cdot e^{2x} = -e^{2x}/2 + xe^{2x} + C \, .$$

In general, when the linear equation

$$y' + a(x)y = b(x)$$

is multiplied by some factor F , we have

$$y'F(x) + a(x)yF(x) = b(x)F(x) \, .$$

Then, the approach via the Product Rule applies when that equation is identical to this equation:

$$y'F(x) + yF'(x) = b(x)F(x) \, ,$$

i.e., when F satisfies:

$$F' = a(x)F \, .$$

This ODE always has a solution according to the last theorem:

$$F = e^{\int a(x) \, dx} \, ,$$

as long as function a is integrable. Function F defined this way is called the *integrating factor* of the linear equation.

Theorem 1.7.4: Non-homogeneous Linear ODE

Every solution of the ODE

$$y' + a(x)y = b(x)$$

is given by:

$$y = e^{-A(x)} \int b(x)e^{A(x)} \, dx$$

where A is any antiderivative of a :

$$A(x) = \int a(x) \, dx$$

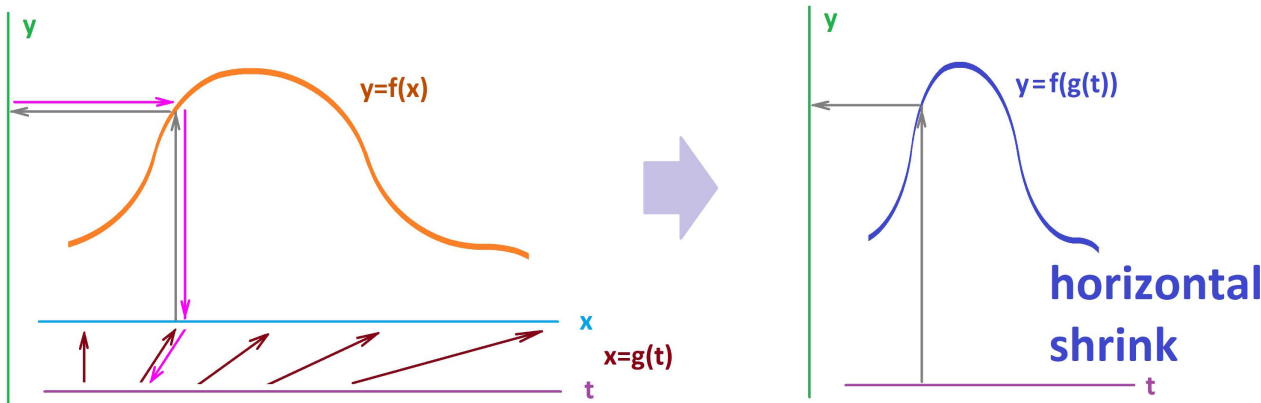
1.8. Change of variables in ODEs

Change of variables is a common tool in sciences and mathematics. The idea of finding a new variable that *might* make the problem simpler is especially pervasive in mathematics. Some simple examples are:

- change of units (seconds to minutes, meters to feet, Celsius to Fahrenheit etc.)

- change of frame of reference (from the Earth to the Sun as the center in order to calculate the motion of the planet, then back to the Earth as the center to calculate the motion of a spacecraft heading for the Moon, etc.)
- representing a function as a composition

The first approach is to try to change the *independent* variable. The result is a horizontal transformation of the plane and the graphs on it:



Example 1.8.1: integration by substitution

Let’s review a simple example of *integration by substitution*:

$$\int x e^{x^2} \, dx = ?$$

Recognizing the presence of a *composition*:

$$y = e^{x^2} \, ,$$

we split it. We introduce (insert!) an intermediate variable, u :

$$x \mapsto u \mapsto y \, ,$$

with

$$u = x^2 \, .$$

Then, the differential of u is this differential form:

$$du = 2x \, dx \, .$$

We substitute these two into the integral:

$$\int x e^{x^2} \, dx = \int x e^{x^2} \Big|_{x^2=u} \, dx \Big|_{dx=\frac{du}{2x}} = \int x e^u \frac{du}{2x} = \frac{1}{2} \int e^u \, du \, .$$

At this point,

- *change of variables* is complete, and
- the new integral is simpler!

Exercise 1.8.2

Carry out the substitution $v = x^3$ in the above integral. Hint: Yes, third power.

Example 1.8.3: substitution in ODEs

Let’s recast the above construction in terms of differential equations.

The integral may be seen as the answer... to what question? An ODE. The ODE corresponding to the integral is:

$$y' = xe^{x^2}.$$

Let’s use the Leibniz notation:

$$\frac{dy}{dx} = xe^{x^2}.$$

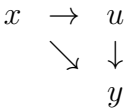
We have *two variables* related via an unknown function: $y = y(x)$. We introduce a new variable:

$$u = x^2.$$

We have *three variables* related to each other via:

- an unknown function: $y = y(x)$ that is still to be found,
- the change of variables function: $u = x^2$, and
- another unknown function: $y = y(u)$ to be found first.

The dependence diagram is below:



The three relations between the three variables have their derivatives:

original:	$x \longrightarrow \longrightarrow \longrightarrow y$ $\frac{dy}{dx} = xe^{x^2}$
decomposed:	$x \longrightarrow u \longrightarrow y$ $\frac{du}{dx} = 2x \qquad \frac{dy}{du} = ?$

Two are known and one is unknown. Fortunately, these derivatives are connected via the *Chain Rule*:

$$\frac{du}{dx} \cdot \frac{dy}{du} = \frac{dy}{dx}.$$

Therefore, our ODE becomes the following after this substitution:

$$\frac{dy}{dx} = xe^{x^2} \implies \frac{du}{dx} \cdot \frac{dy}{du} = xe^u \implies 2x \frac{dy}{du} = xe^u \implies \frac{dy}{du} = \frac{1}{2}e^u.$$

We have accomplished the following:

- The *change of variables* is complete.
- The new ODE is simpler!

This is the summary:

$$\frac{dy}{dx} = xe^{x^2} \longrightarrow \frac{dy}{du} = \frac{dy}{du}.$$

We solve the new ODE (with respect to u) easily:

$$y = \int \frac{1}{2}e^u du = \frac{1}{2}e^u + C.$$

Since the variable u was made up, we need the *back-substitution*, $u = x^2$, giving us the solution to the original ODE (with respect to x):

$$y = \frac{1}{2}e^{x^2} + C.$$

Exercise 1.8.4

Carry out the substitution $v = x^3$ in the above ODE.

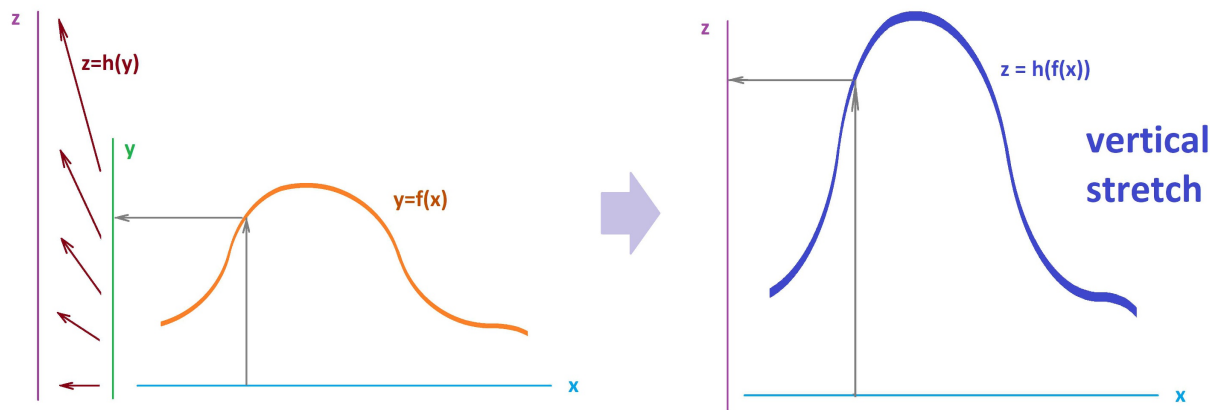
Exercise 1.8.5

Carry out the substitution $v = \sin x$ in the ODE $y' = \cos^3 x$.

Exercise 1.8.6

Execute the substitution

The second approach is to try to change the *dependent* variable. The result is a vertical transformation of the plane and the graphs on it:



The simplest such substitution is the *shift*:

$$y = z + a ,$$

where z is the new dependent variable. If we want to concentrate on a single point $y = a$ (and its vicinity) at a time, the shift allows us to move that point to zero.

Example 1.8.7: migration

For example, suppose we have a *linear ODE*:

$$y' = my + b, \quad m \neq 0 .$$

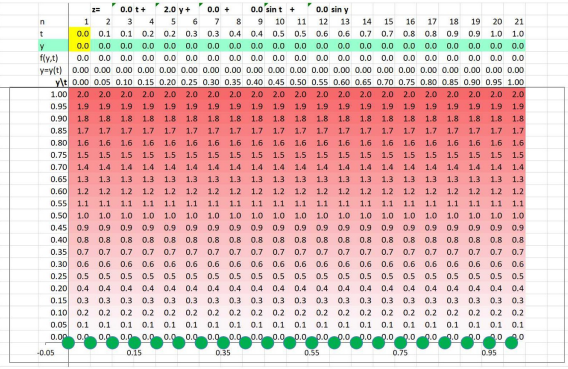
It can represent population growth/decay accompanied by *migration* at a constant rate. The ODE is close to $y' = my$ (no migration) that we know how to handle. But how do we get there? Let's investigate.

Let's compare the corresponding difference equations:

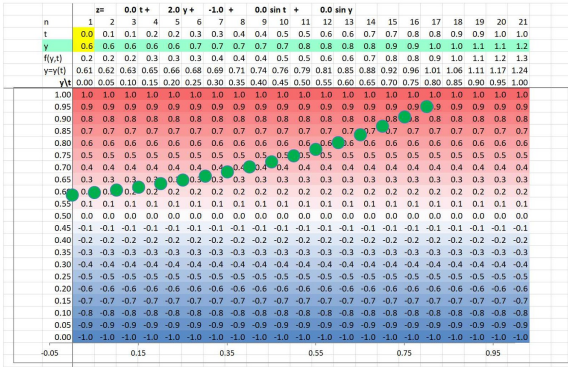
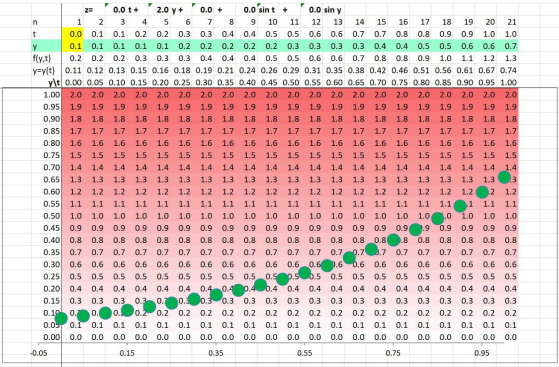
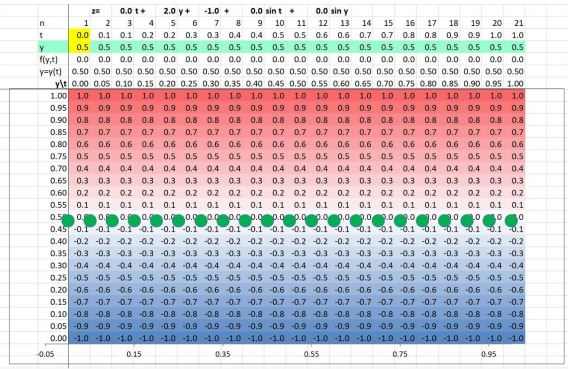
$$\frac{\Delta y}{\Delta x} = my \quad \text{vs.} \quad \frac{\Delta y}{\Delta x} = my + b .$$

We plot a couple of their solutions:

natural growth



natural growth with outflux



In comparison to the standard exponential growth (left), there seems to be a vertical shift (right). But how far?

Let's choose a new variable:

$$z = y - a.$$

That's a vertical shift (with a still to be found). Then, first,

$$y' = (z + a)' = z',$$

and, second,

$$my + b = m(z + a) + b = mz + ma + b = mz + m \left(a + \frac{b}{m} \right).$$

For the last term to disappear we just select:

$$a = -\frac{b}{m}.$$

We have a new ODE with respect to z :

$$z' = mz.$$

We have accomplished the following:

- The *change of variables* is complete.
- The new ODE is simpler!

This is the summary:

$$\frac{dy}{dx} = my + b \longrightarrow \frac{dz}{dx} = mz.$$

We solve it as before:

$$z = Ce^{mt}.$$

After a back-substitution, we have for the original ODE:

$$y - a = Ce^{mt},$$

or

$$y = -\frac{b}{m} + Ce^{mt}.$$

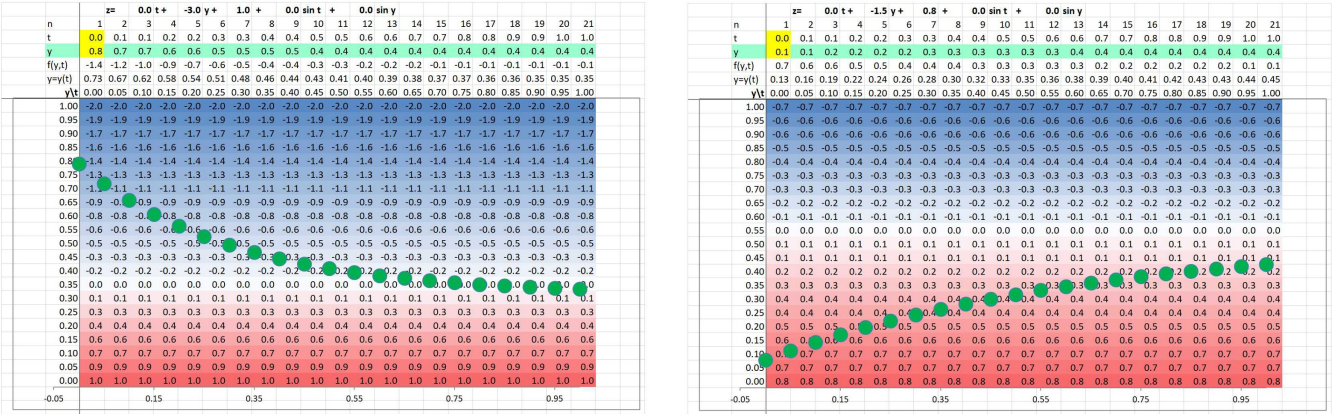
The whole picture just shifts up!

Example 1.8.8: banking

What happens if we combine regular deposits with compounded interest?

According to the last example, the same exponential growth as in the case without deposits but shifted down (not up!). What happens if we combine regular withdrawals with compounded interest? The same exponential growth as in the case without withdrawals but shifted up.

Alternatively, what if our investment is losing value but we continue to put money in it? The exponential decay or a stagnating growth toward a non-zero value:

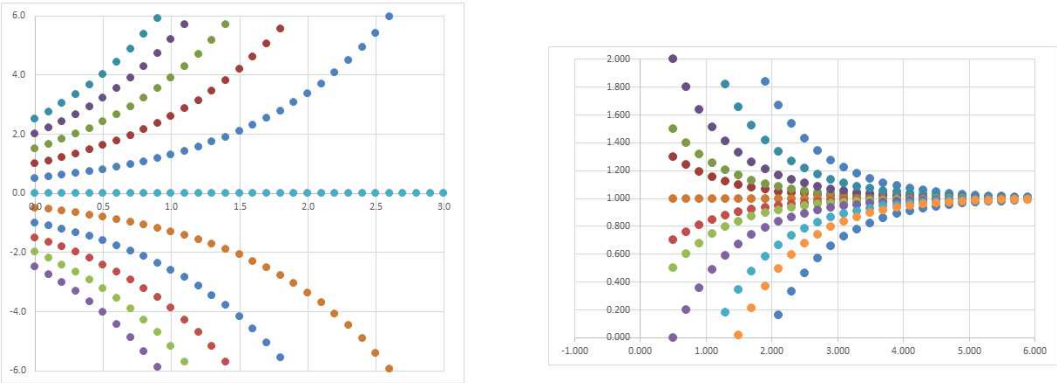


Example 1.8.9: Newton's Law of Cooling

Let's write the ODEs of the population growth model and the cooling law model next to each other:

$y' = ky$ and $y' = k \cdot (r - y)$.

They seem similar! Let's plot the solutions of the corresponding difference equations of these models:



They seem similar!

The comparison suggests a substitution: a vertical shift and a horizontal flip. We execute one at a time. The shift is given by:

$z = y - r$.

Then the inverse substitution is:

$y = z + r$.

Then this is what happens to our ODE as we substitute:

$\frac{dy}{dt} = k(r - y) \implies \frac{d(z + r)}{dt} = k(r - (z + r)) \implies \frac{dz}{dt} = -kz$.

Except for the minus sign in the beginning, this is the population ODE! Now the flip:

$s = -t$.

Then the inverse substitution is:

$$t = -s.$$

Now the derivatives:

$$\frac{dz}{dt} = \frac{dz}{ds} \frac{ds}{dt} = -\frac{dz}{ds}.$$

Our ODE becomes:

$$\frac{dz}{ds} = kz.$$

We have confirmed that the solution set of the cooling ODE is that of the population ODE sifted vertically by r and then flipped horizontally!

Furthermore, we have the explicit solutions too via a back-substitutions:

$$z = Ce^{ks} \implies y - r = Ce^{k(-t)} \implies y = r + Ce^{-kt}.$$

Exercise 1.8.10

Execute the substitution

1.9. Euler's method: back to the discrete

Unfortunately, an ODE typically has no algebraic solution!

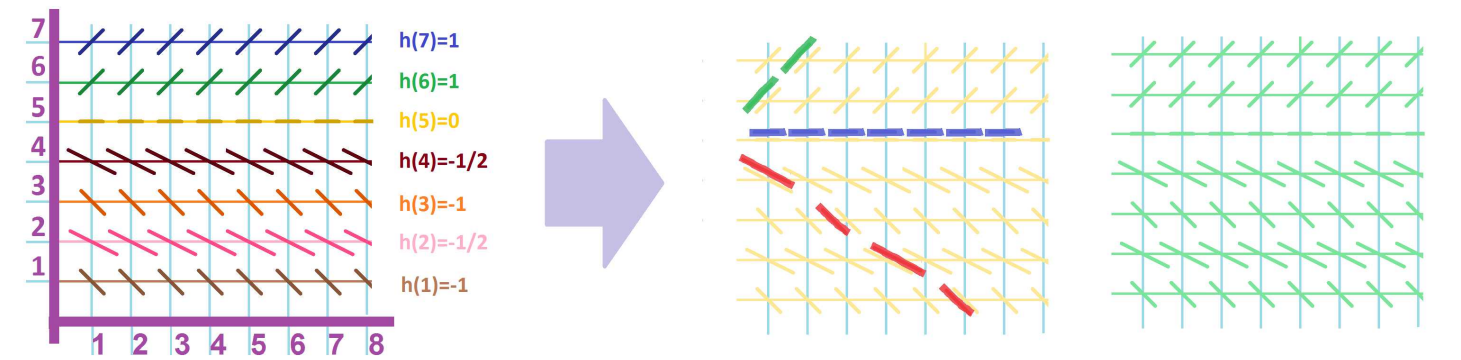
We would like to approximate solutions of a general IVP:

$$y' = f(t, y), \quad y(t_0) = y_0.$$

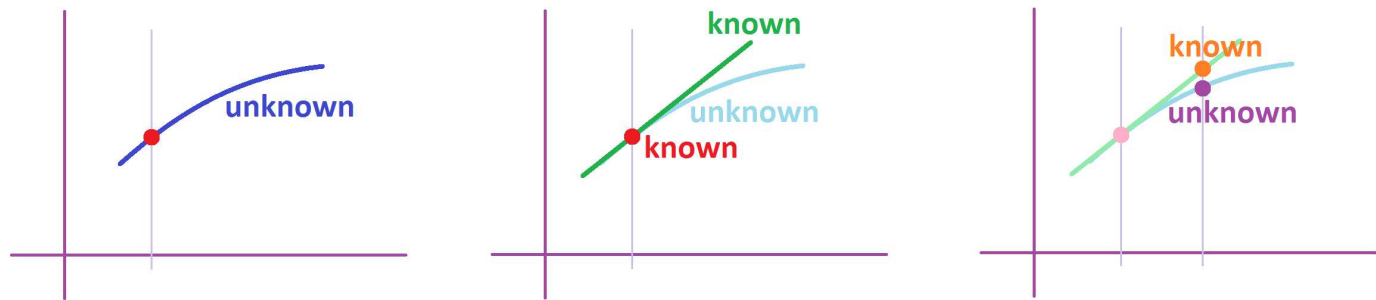
The IVP tells us:

- where we are (the initial condition), and
- the direction we are going (the ODE).

It is then easy to produce a solution if we assume that the change is incremental: as we arrive to a new location, we, once again, know where we are and where we are going.

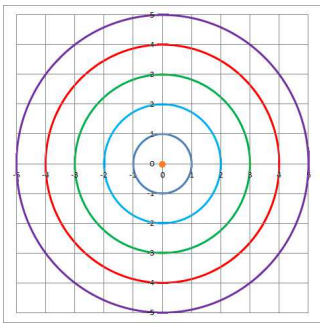


The method presented here follows the idea from [Chapter 2DC-6](#): the unknown solution is replaced with its *best linear approximation*.



Example 1.9.1: family of circles

Let’s consider again these concentric circles:



They are the solutions of the ODE:

$$y' = f(x,y) = -\frac{x}{y}.$$

We will be solving numerous initial value problems while staying away from the x -axis where there are no solutions.

We choose the increment of x :

$$\Delta x = 1.$$

We start with this initial condition:

$$x_0 = 0, \quad y_0 = 1.$$

We substitute these two numbers into the equation:

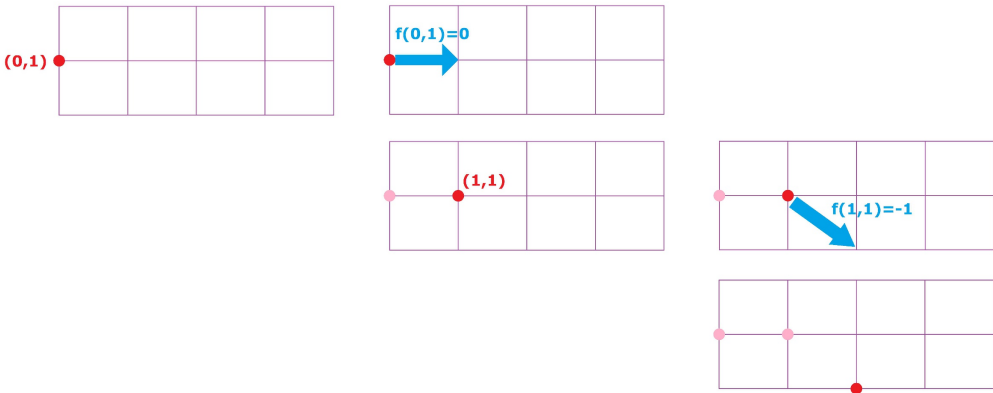
$$y' = -\frac{0}{1} = 0.$$

This is the slope of the line we will follow. How far? As far as the increment of x allows. Therefore, the increment of y is

$$\Delta y = 0 \cdot \Delta x = 0 \times 1 = 0.$$

Our next location on the xy -plane is then:

$$x_1 = x_0 + \Delta x = 0 + 1 = 1, \quad y_1 = y_0 + \Delta y = 1 + 0 = 1.$$



This computation gives us a *new initial condition*:

$$x_1 = 1, \quad y_1 = 1.$$

We again substitute these two numbers into the equation:

$$y' = -\frac{1}{1} = -1.$$

This is the slope of the line we will follow. Therefore, the increment of y is

$$\Delta y = -1 \cdot \Delta x = -1.$$

Our next location on the xy -plane is then:

$$x_2 = x_1 + \Delta x = 1 + 1 = 2, \quad y_2 = y_1 + \Delta y = 1 + (-1) = 0.$$

We have ended up on the x -axis and stop. These three points form an approximate solution.

We start all over from another initial condition:

$$x_0 = 0, \quad y_0 = 2.$$

We substitute these two numbers into the equation:

$$y' = -\frac{0}{2} = 0,$$

producing the slope we will follow. The increment of y is

$$\Delta y = 0 \cdot \Delta x = 0 \cdot 1 = 0.$$

Our next location on the xy -plane is then:

$$x_1 = x_0 + \Delta x = 0 + 1 = 1, \quad y_1 = y_0 + \Delta y = 2 + 0 = 2.$$

A new initial condition appears:

$$x_0 = 1, \quad y_0 = 2.$$

We again substitute these two numbers into the equation:

$$y' = -\frac{1}{2} = -1/2,$$

producing the slope to follow. The increment of y is

$$\Delta y = -1/2 \cdot \Delta x = -1/2.$$

Our next location on the xy -plane is then:

$$x_2 = x_1 + \Delta x = 1 + 1 = 2, \quad y_2 = y_1 + \Delta y = 2 + (-1/2) = 3/2.$$

One more IVP:

$$x_2 = 2, \quad y_2 = 3/2.$$

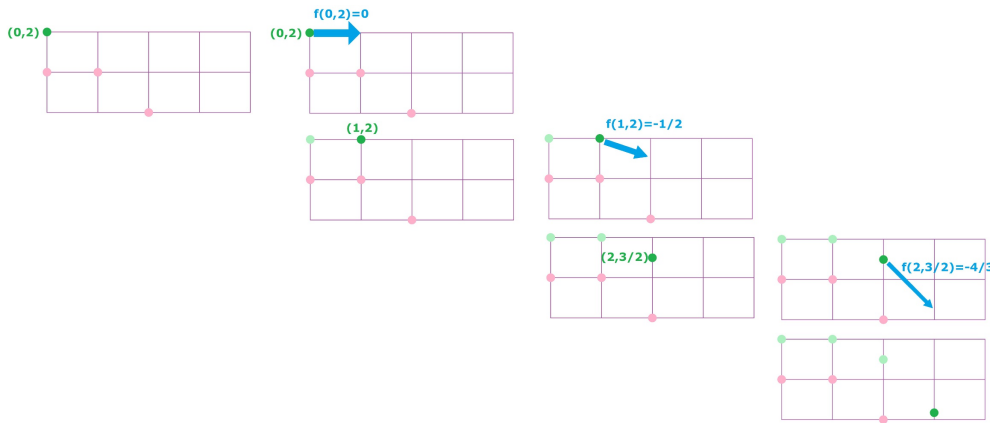
The increment of y is

$$\Delta y = -\frac{x}{y} \cdot \Delta x = -\frac{2}{3/2} \cdot 1 = -4/3.$$

Our next location on the xy -plane is then:

$$x_3 = x_2 + \Delta x = 2 + 1 = 3, \quad y_3 = y_2 + \Delta y = 3/2 - 4/3 = 1/6.$$

We would have to pass the x -axis with next step and we stop now. These four points, in addition to the previous three, form a very crude approximation of our circular solutions:



In fact, we see what isn’t supposed to happen: the solution goes below the x -axis!

From the point of view of motion, this is the method’s interpretation:

- At our current location and current time, we examine the ODE to find the velocity and then move with this velocity to the next location.

Definition 1.9.2: Euler solution

The *Euler solution* with increment $h > 0$ of the IVP

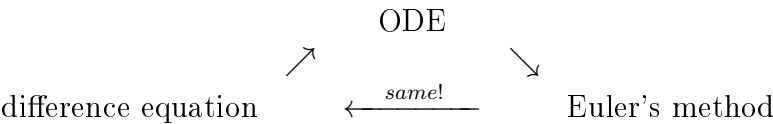
$$y' = f(t, y), \quad y(t_0) = y_0,$$

is the sequence $\{y_n\}$ of real numbers given by:

$$y_{n+1} = y_n + f(t_n, y_n) \cdot h$$

where $t_{n+1} = t_n + h$.

The most important fact to know about Euler’s method is that if we derived our ODE from a discrete model – and from a difference equation (discrete ODE) via $\Delta t \rightarrow 0$ – Euler’s method will bring us right back to it:



Exercise 1.9.3

Execute the method for the above example and $h = 1/2$.

Example 1.9.4: circles

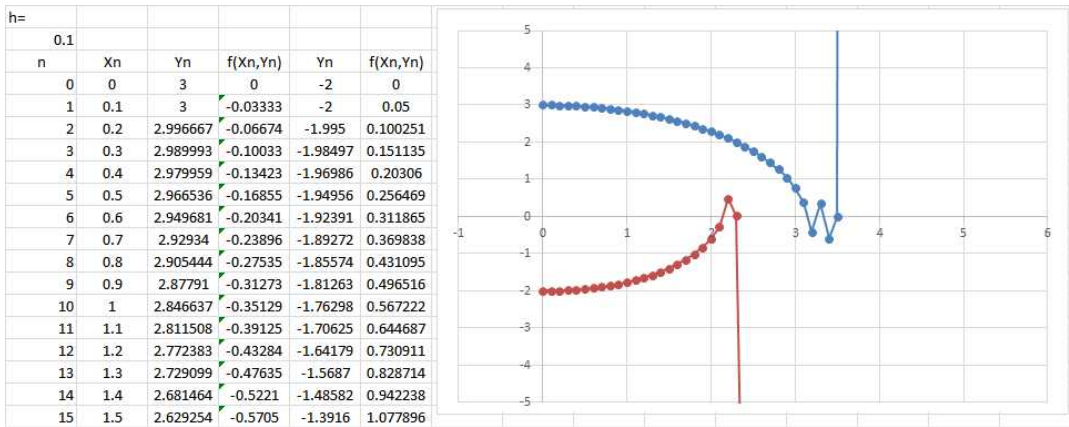
Let’s now carry out this procedure with a spreadsheet for the ODE:

$$y' = -\frac{y}{x}.$$

The formula for y_n is:

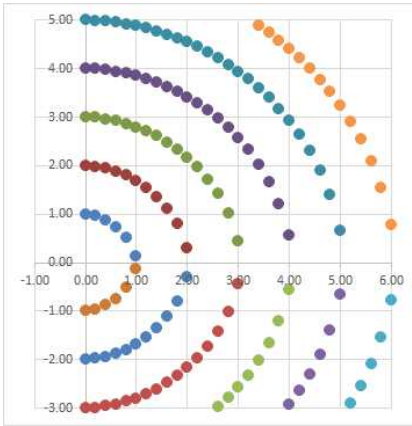
$$=R[-1]C+R[-1]C[1]*R2C1$$

The results are less crude:



However, the approximations appear to behave erratically when they get close to the x -axis. The reason is the division by y which may be very small. In fact the solutions are supposed to stop at the x -axis, but, by design of Euler's method, they can't.

Elsewhere, the Euler solutions are close to perfect:



The separation of variables solution:

$$\frac{dy}{dx} = -\frac{x}{y} \implies ydy = -xdx \implies \int ydy = -\int xdx \implies \frac{y^2}{2} = -\frac{x^2}{2} + C.$$

Exercise 1.9.5

Finish the solution.

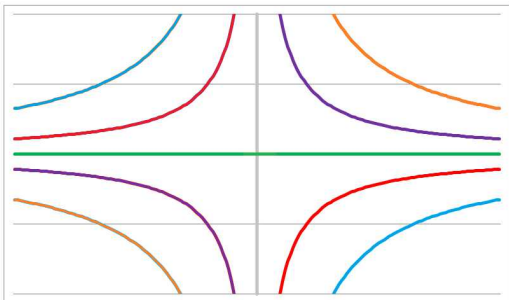
Warning!

The Euler solutions might be unaffected by the non-existence property of the ODE.

Example 1.9.6: hyperbolas

Let's consider again these hyperbolas:

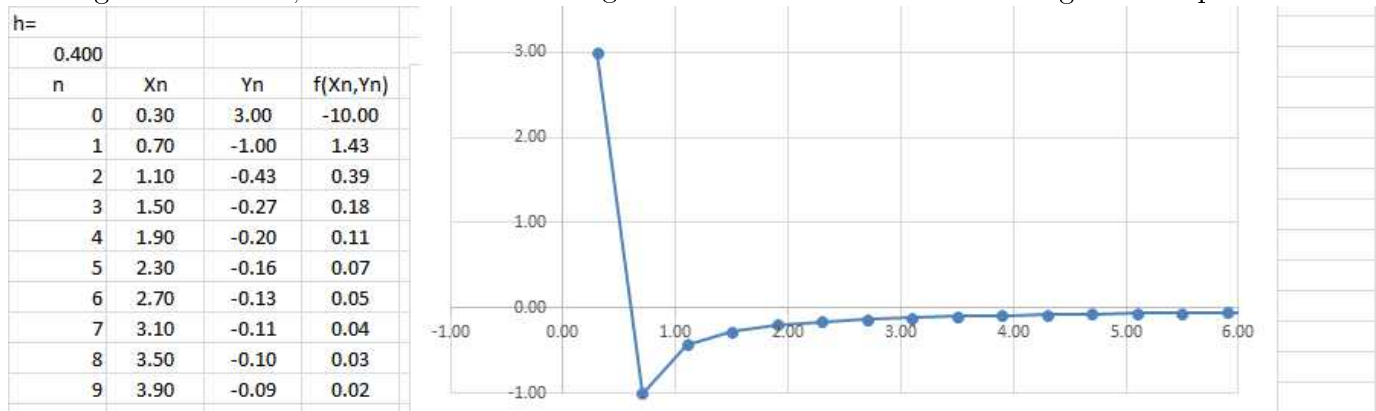
$$xy = C.$$



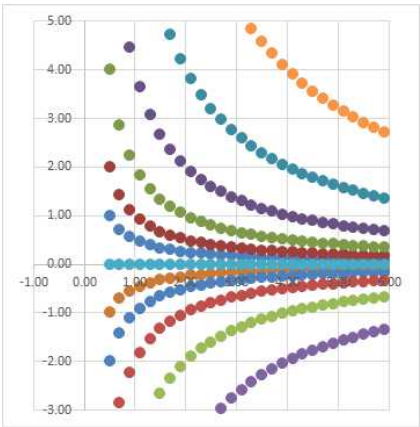
They are the solutions of the ODE:

$$y' = \frac{y}{x}.$$

For larger values of h , the Euler solution might cross the x -axis demonstrating non-uniqueness:



Several better looking Euler solutions are shown below:



Even then, is this asymptotic convergence toward the x -axis or do they merge?

The separation of variables solution:

$$\frac{dy}{dx} = -\frac{y}{x} \implies \frac{dy}{y} = -\frac{dx}{x} \implies \int \frac{dy}{y} = -\int \frac{dx}{x} \implies \ln y = -\ln x + K.$$

Exercise 1.9.7

Finish the solution.

Warning!

Euler’s method might introduce non-uniqueness to an ODE.

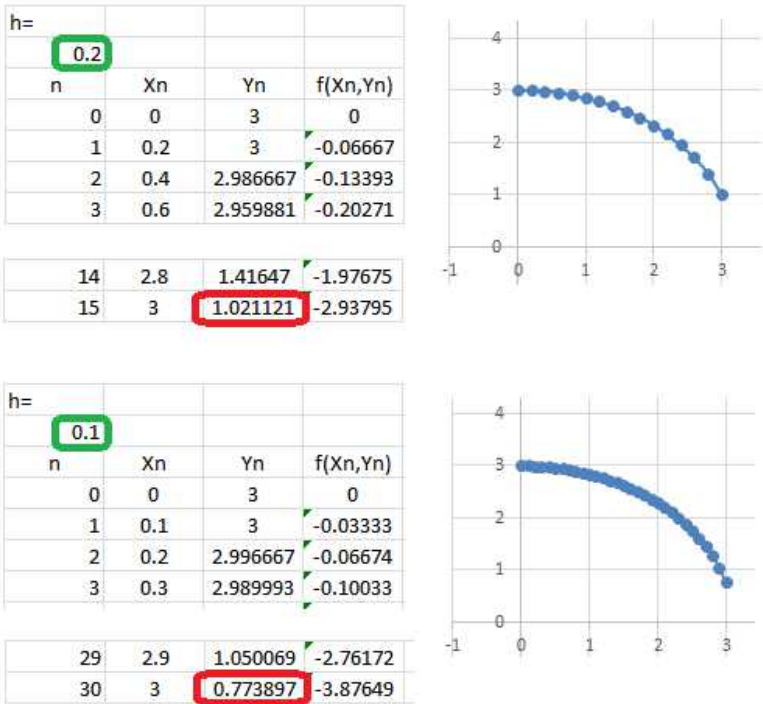
1.10. How large is the difference between the discrete and the continuous?

Example 1.10.1: approximations?

How far apart are the solutions of a pair of matching discrete and continuous ODEs? In other words, how far is the Euler solution from the actual solution of the ODE for the derivatives? For example, consider:

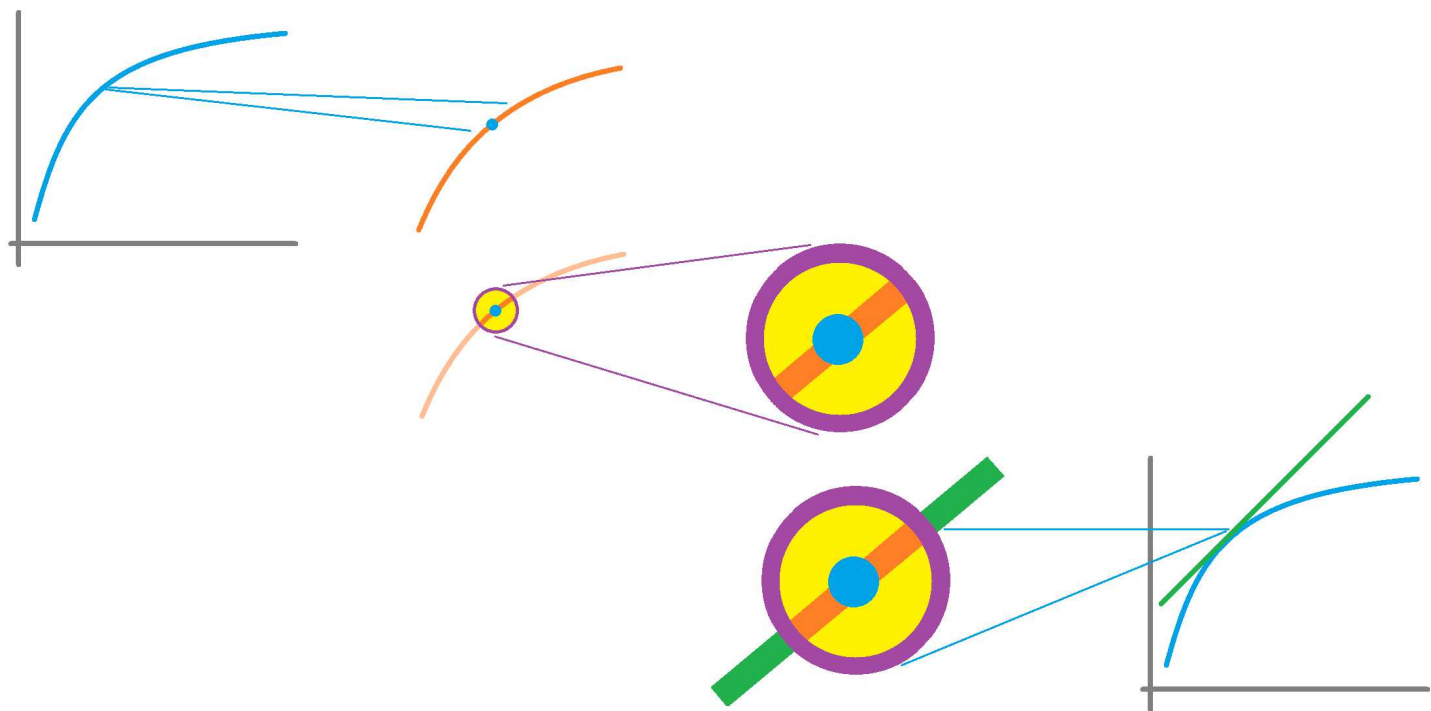
$$\frac{\Delta y}{\Delta t} = -\frac{t}{y} \text{ and } y' = -\frac{t}{y}.$$

The solutions of the latter are circles. So, if one such solution starts at $(0, y_0)$, i.e., $t = 0$ and $y = y_0$, then the curve is supposed to end at the same location on the t -axis, i.e., $(y_0, 0)$. This is not what we see when we lot a solution of the former equation:



Why does this happen? The solutions of the former decrease slower than those of the latter. Of course, the difference diminishes with the value of $h = \Delta t$.

Approximating functions is like approximating numbers (such as π or the Riemann integral) but harder. Recall from Chapter 2DC-6 that *linearization* means replacing a given function $y = f(x)$ with a linear function $y = L(x)$ that best approximates it at a given point. It is called its best linear approximation and its happens to be the linear function the graph of which is the tangent line at the point. The replacement is justified by the fact that when you zoom in on the point, the tangent line will merge with the graph:



However, there is a more basic approximation: a constant function, $y = C(x)$.

Example 1.10.2: square root

Let’s review this example from [Chapter 1PC-2C-6](#): how do we compute $\sqrt{4.1}$ without actually evaluating $f(x) = \sqrt{x}$? We approximate. And to approximate the number of $\sqrt{4.1}$, we approximate the function $f(x) = \sqrt{x}$ “around” $a = 4$.

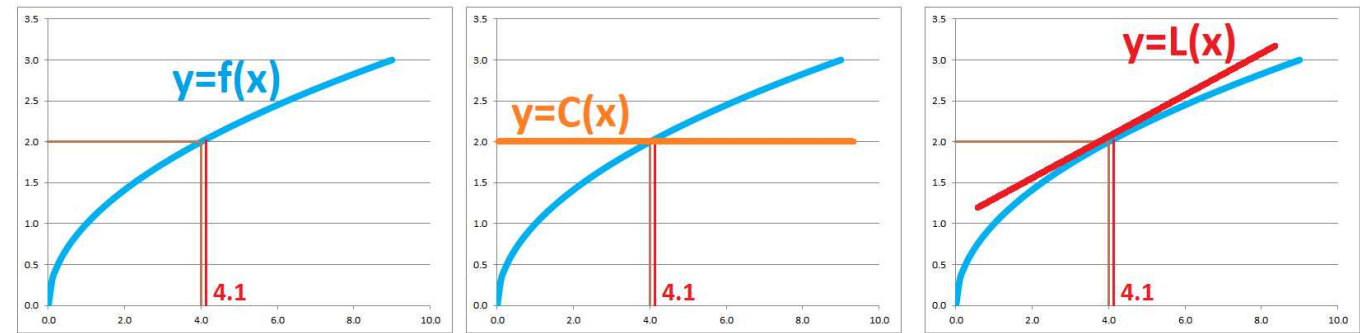
We first approximate the function with a *constant* function:

$$C(x) = 2.$$

This value is chosen because $f(a) = \sqrt{4} = 2$. Then we have:

$$\sqrt{4.1} = f(4.1) \approx C(4.1) = 2.$$

It is a crude approximation:



The other, linear, approximation is visibly better. We approximate the function with a *linear* function:

$$L(x) = 2 + \frac{1}{4}(x - 4).$$

This value is chosen because $f(a) = \sqrt{4} = 2$ and $f'(a) = \frac{1}{4}$. Then we have:

$$\sqrt{4.1} = f(4.1) \approx L(4.1) = 2 + \frac{1}{4}(4.1 - 4) = 2.025.$$

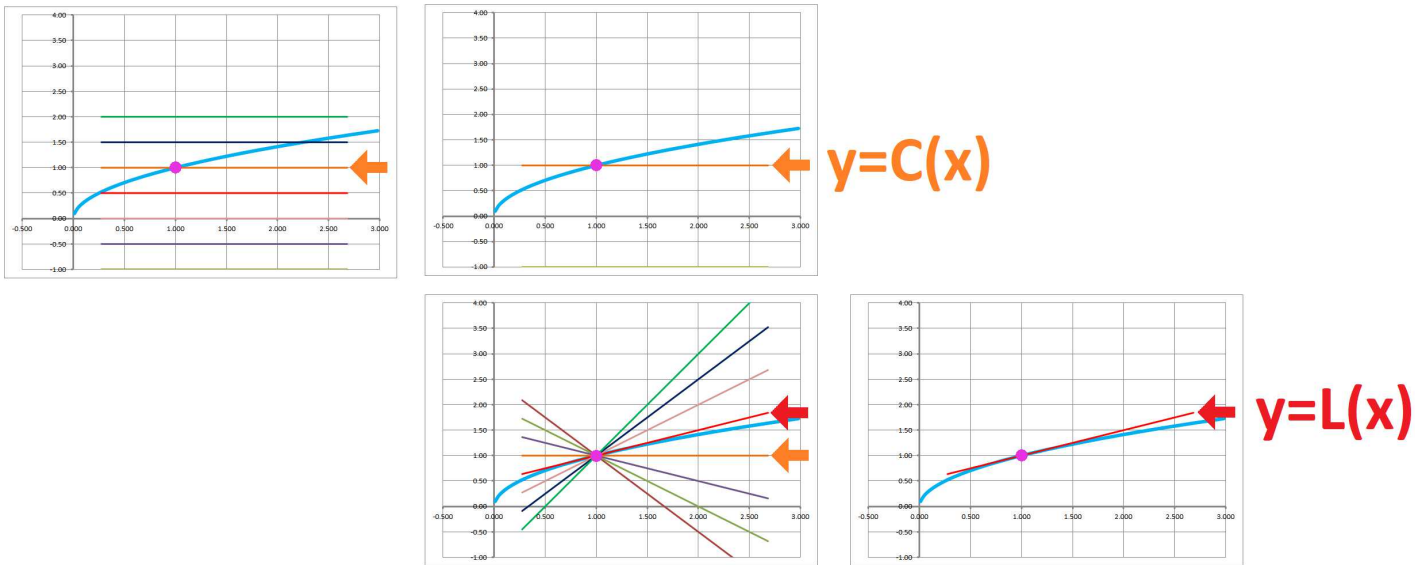
We have for a function $y = f(x)$ and $x = a$:

A constant approximation: $C(x) = f(a)$

A linear approximation: $L(x) = f(a) + f'(a)(x - a)$

We should notice early on that the latter just adds a new (linear) term to the former! Also, the latter is better than the former – but only when we need more accuracy. Otherwise, the former is worse because it requires more computation.

Below we illustrate how we attempt to approximate a function around the point $(1, 1)$ with *constant* functions first; from those we choose the horizontal line through the point. This line then becomes one of many *linear* approximations of the curve that pass through the point; from those we choose the tangent line.



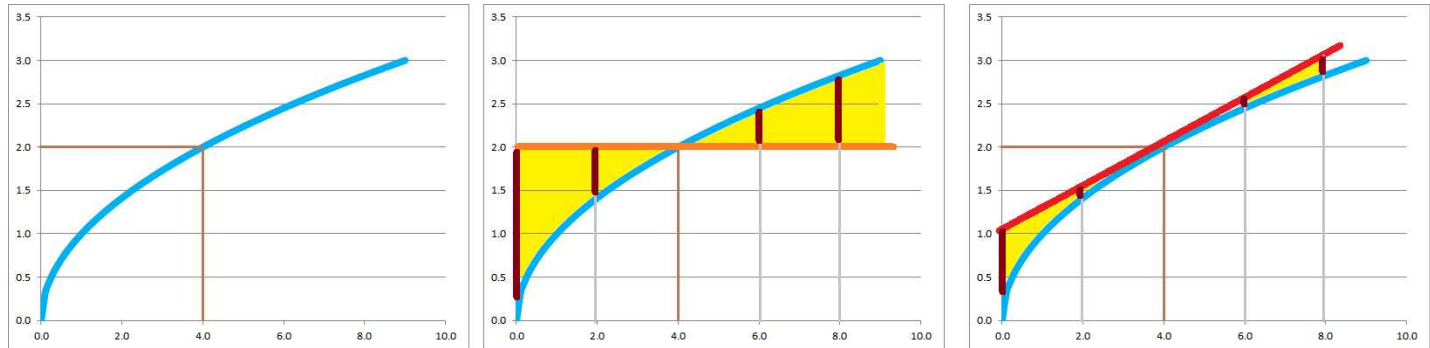
Now, we shall see that these are just the two first steps in a *sequence* of approximations! The tangent line becomes one of many *quadratic* curves – parabolas – that pass through the point... and are tangent to the curve. Which one of *those* do we choose?

In order to answer that, we need to review and understand how the best constant and the best linear approximations were chosen. In what way are they the best?

Suppose a function $y = f(x)$ is given and we would like to approximate its behavior in the vicinity of a point, $x = a$, with another function $y = T(x)$. The latter is to be taken from some class of functions that we find suitable.

What we need to consider is the *error*, i.e., the difference between the function f and its approximation T :

$$E(x) = |f(x) - T(x)|.$$



We are supposed to minimize the error function in some way.

Of course, the error function $y = E(x)$ is likely to grow with no limit as we move away from our point of interest, $x = a$... but we don't care. We want to minimize the difference in the *vicinity* of a which means making sure that the limit of the error as $x \rightarrow a$ goes to 0!

Theorem 1.10.3: Best constant approximation

Suppose f is continuous at $x = a$ and

$$C(x) = k$$

is any of its constant approximations (i.e., arbitrary constant functions). Then, the error E of the approximation approaches 0 at $x = a$ if and only if the constant is equal to the value of the function f at $x = a$; i.e.,

$$\lim_{x \rightarrow a} (f(x) - C(x)) = 0 \iff k = f(a)$$

That's the analog of the following theorem from [Chapter 2DC-6](#).

Theorem 1.10.4: Best linear approximation

Suppose f is differentiable at $x = a$ and

$$L(x) = f(a) + m(x - a)$$

is any of its linear approximations. Then, the error E of the approximation approaches 0 at $x = a$ faster than $x - a$ if and only if the coefficient of the linear term is equal to the value of the derivative of the function f at $x = a$; i.e.,

$$\lim_{x \rightarrow a} \frac{f(x) - L(x)}{x - a} = 0 \iff m = f'(a)$$

Comparing these two conditions:

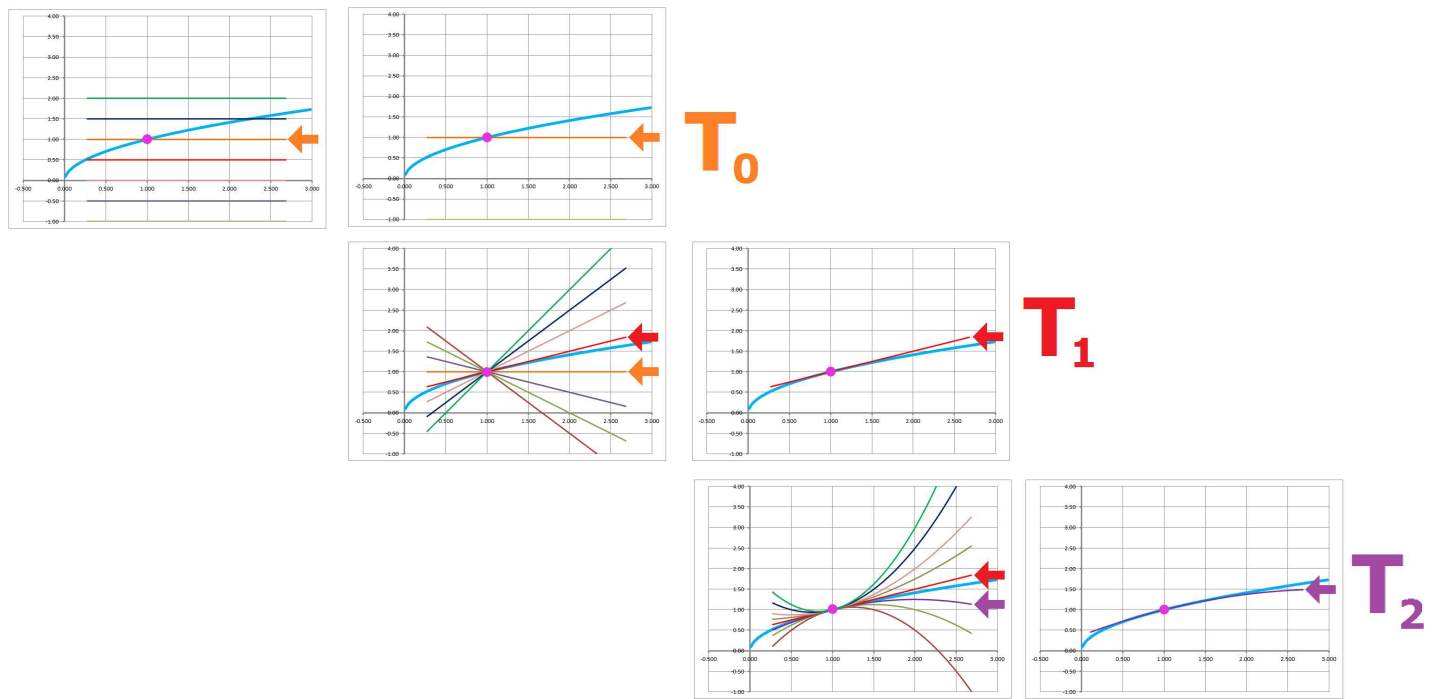
$$f(x) - C(x) \rightarrow 0 \text{ and } \frac{f(x) - L(x)}{x - a} \rightarrow 0,$$

reveals the similarity and the difference in how we minimize the error! The difference is in the degree: *how fast the error function goes to zero*. Indeed, we learned in [Chapter 2DC-6](#) that the latter condition means that $f(x) - L(x)$ converges to 0 faster than $x - a$, i.e.,

$$f(x) - L(x) = o(x - a),$$

there is no such restriction for the former.

So far, this is what we have discovered: linear approximations are built from the best constant approximation by adding a linear term. The best one of those has the slope (its own derivative) equal to the derivative of f at a . How the sequence of approximations will progress is now clearer: quadratic approximations are built from the best linear approximation by adding a quadratic term.



But which one of those is the best?

Theorem 1.10.5: Best quadratic approximation

Suppose f is twice continuously differentiable at $x = a$ and

$$Q(x) = f(a) + f'(a)(x - a) + p(x - a)^2$$

is any of its quadratic approximations. Then, the error E of the approximation approaches 0 at $x = a$ faster than $(x - a)^2$ if and only if the coefficient of the quadratic term is equal to half of the value of the second derivative of the function f at $x = a$; i.e.,

$$\lim_{x \rightarrow a} \frac{f(x) - Q(x)}{(x - a)^2} = 0 \iff p = \frac{1}{2}f''(a)$$

Once again, the condition of the theorem means that $f(x) - Q(x)$ converges to 0 faster than $(x - a)^2$, or

$$f(x) - Q(x) = o((x - a)^2).$$

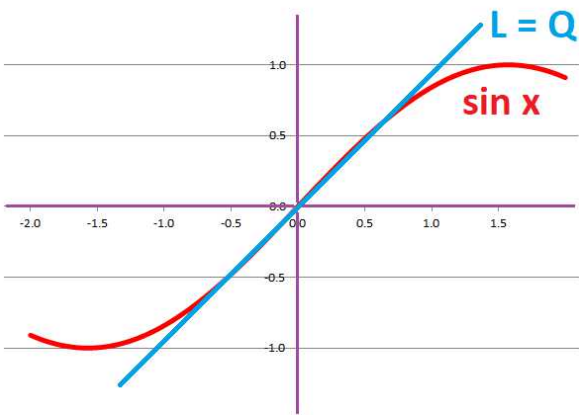
We start to see a pattern:

- The degrees of the approximating polynomials are growing.
- The degrees of the derivatives being taken into account are growing too.

Example 1.10.6: sine

Let's approximate $f(x) = \sin x$ at $x = 0$. First, the values of the function and the derivatives:

$$\begin{aligned} f(x) &= \sin x &\implies f(0) &= 0 &\implies C(x) &= 0 \\ f'(x) &= \cos x &\implies f'(0) &= 1 &\implies L(x) &= x \\ f''(x) &= -\sin x &\implies f''(0) &= 0 &\implies Q(x) &= ? \end{aligned}$$



Therefore, the best quadratic approximation is:

$$Q(x) = 0 + 1(x - 0) - \frac{0}{2}(x - 0)^2 = x.$$

Same as the linear! Why? Because sin is odd.

We used sequences of *numbers* to approximate other numbers (Chapter 5); now we will use sequences of *functions* to approximate other functions. In order to be able to go beyond quadratic in our sequence of polynomial approximations, we rename them according to their *degrees*:

$$\begin{aligned} T_0(x) &= C(x), \\ T_1(x) &= L(x), \\ T_2(x) &= Q(x), \\ &\dots \end{aligned}$$

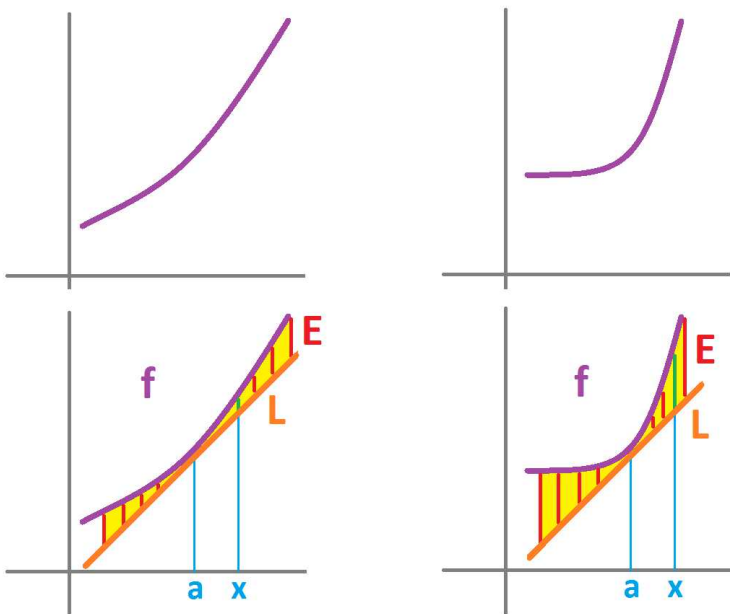
Example 1.10.7: square root

Back to the original example of $f(x) = \sqrt{x}$ at $a = 4$. One can guess where this is going:

constant:		$2 = T_0(x)$	$f - T_0 = o(1)$
linear:		$\frac{1}{4}(x - 4) + 2 = T_1(x)$	$f - T_1 = o(x - a)$
quadratic:	$-\frac{1}{2 \cdot 32}(x - 4)^2$	$+\frac{1}{4}(x - 4) + 2 = T_2(x)$	$f - T_2 = o((x - a)^2)$
cubic:	$(?)(x - 4)^3$	$-\frac{1}{2 \cdot 32}(x - 4)^2 + \frac{1}{4}(x - 4) + 2 = T_3(x)$	$f - T_3 = o((x - a)^3)$
	\vdots	\vdots	\vdots

We add a term every time we move down to the next degree; it's a *recursion*! Such a sequence (a sequence of functions!) is called a “series”.

We consider a solution of a difference equation (discrete ODE) and that of the corresponding continuous ODE with the same initial condition, $y(t_0) = y_0$. Their difference of the two is a function and the absolute value of this function evaluated at some $t > t_0$ is referred to as the *error*.



To estimate the error, we consider the *error bound of the best linear approximation* from [Chapter 2DC-6](#). Suppose y is twice differentiable at $t = t_0$ and $L(t) = y(t_0) + y'(t_0)(t - t_0)$ is its best linear approximation at t_0 . Then, the error satisfies:

$$E(t) = |y(t) - L(t)| \leq \frac{1}{2}K(t - t_0)^2,$$

where K is a bound of the second derivative on the interval from t_0 to t :

$$|y''(c)| \leq K \text{ for all } c \text{ in this interval.}$$

Suppose now that y is a solution of the IVP:

$$y' = f(t, y), \quad y(t_0) = y_0.$$

Provided its derivative is bounded as above, the error of a single step of the corresponding difference equation (i.e., Euler method) is bounded by:

$$E(t_0 + h) = |y(t_0 + h) - y_1| \leq \frac{1}{2}Kh^2,$$

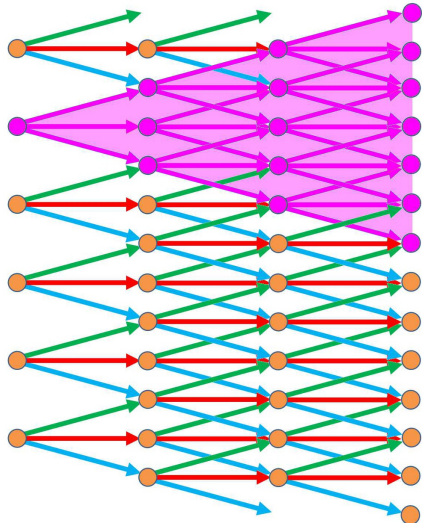
where $y_1 = L_1(t_0 + h)$ is the first step of the approximation that comes from

$$L_1(t) = y_0 + f(t_0, y_0)(t - t_0),$$

the linearization of y at t_0 .

We conclude that *the error of a single step is decreasing quadratically with h* .

Now, let's consider multiple steps. This is what will happen:



What can we do about this *propagation of error*?

If y'' is bounded by the same constant K , we can apply this formula to the second step of the approximation:

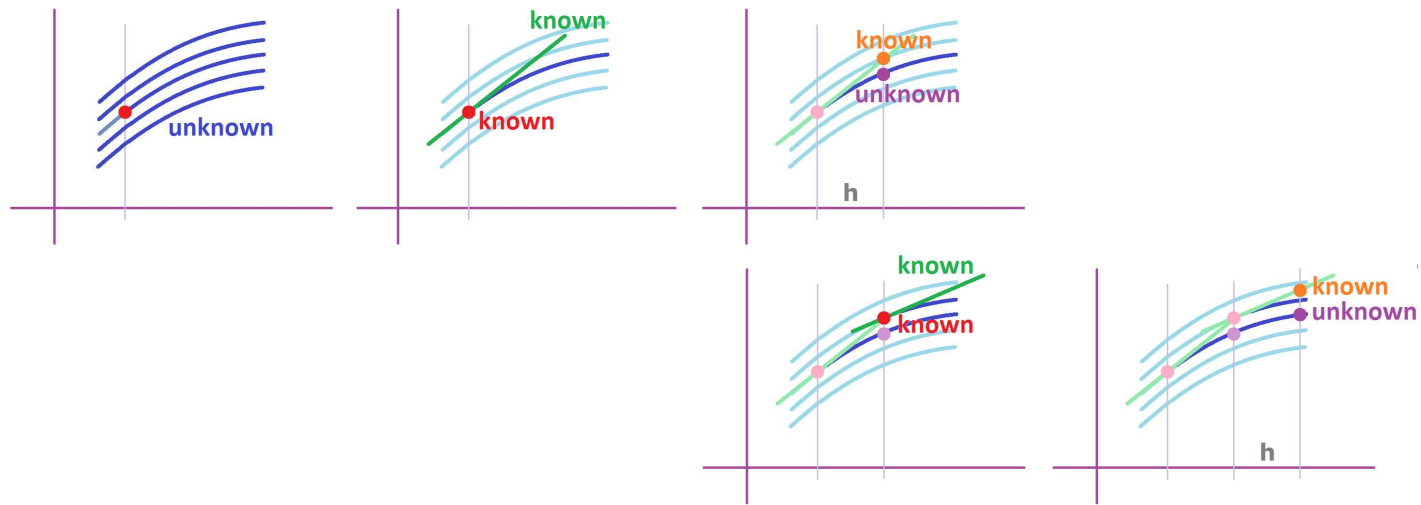
$$|y(t_1 + h) - y_2| \leq \frac{1}{2}Kh^2,$$

where $y_2 = L_2(t_1 + h)$ is the second step of the approximation that comes from

$$L_2(t) = y_1 + f(t_1, y_1)(t - t_1),$$

the linearization of y at t_1 .

Notice that the two error bounds apply to two *different* solutions y of two different IVPs:



Therefore, the best bound for the error of the two steps is the sum of the two (in our case, identical) bounds:

$$E(t_0 + 2h) = |y(t_0 + 2h) - y_2| \leq 2 \cdot \frac{1}{2}Kh^2.$$

And so on... The bound for the error of the n steps is n times the bound:

$$E(t_0 + nh) = |y(t_0 + nh) - y_n| \leq n \cdot \frac{1}{2}Kh^2.$$

Suppose now that we extend the discrete solution with n steps to $t_0 + H$. Then

$$n = \frac{H}{h}.$$

Therefore,

$$E(t_0 + nh) = |y(t_0 + nh) - y_n| \leq \frac{H}{h} \cdot \frac{1}{2}Kh^2 = \frac{1}{2}KHh.$$

We conclude that *the error is decreasing linearly with h .*

Theorem 1.10.8: Error Bound

Suppose a differentiable function $z = f(t, y)$ satisfies

$$\frac{\partial f}{\partial t}(t, y) + \frac{\partial f}{\partial y}(t, y) f(t, y) \leq K,$$

for every (t, y) in the rectangle $t_0 \leq t \leq t_0 + H$, $A \leq y \leq B$. Suppose y is a solution of the IVP

$$y' = f(t, y), \quad y(t_0) = y_0,$$

with the domain $[t_0, t_0 + H]$ and the range within $[A, B]$. Suppose y_n is the n steps of the solution of the discrete counterpart of this IVP with $h = H/n$. Then

their difference satisfies:

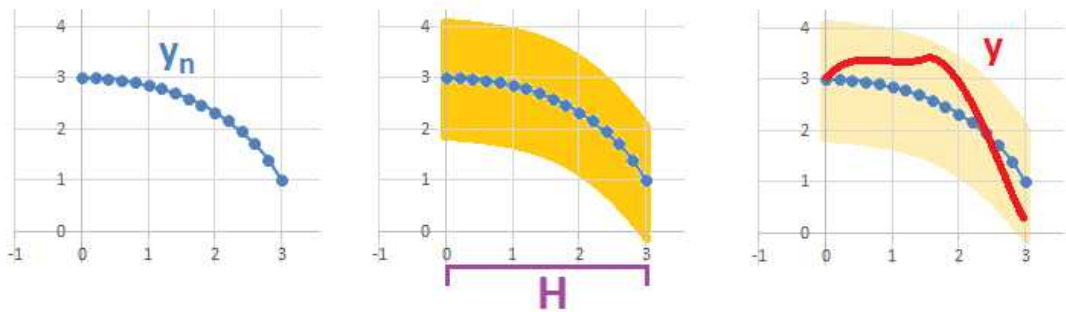
$$|y(t_0 + nh) - y_n| \leq \frac{1}{2}KHh.$$

Proof.

The proof is above; we just use the *Chain Rule* to derive this:

$$y''(t) = \frac{\partial f}{\partial t}(t, y(t)) + \frac{\partial f}{\partial y}(t, y(t)) f(t, y(t)).$$

For each h , the error bound provides a tunnel, of constant width, around the discrete solution that contains the (unknown) solution of the continuous ODE, and vice versa:



Exercise 1.10.9

Prove the last statement. Hint: you need more than just n points as in the theorem.

Exercise 1.10.10

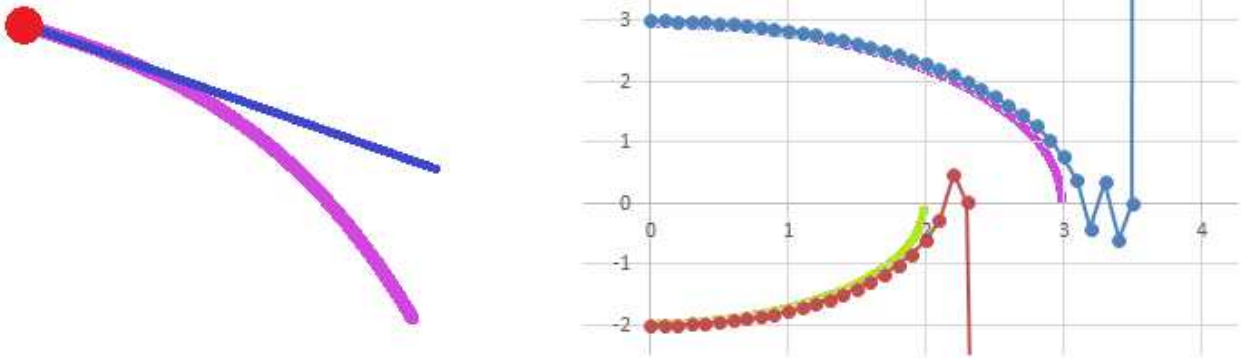
Apply the theorem to the example of concentric circles in the beginning of the section.

Example 1.10.11: approximations of circles

In our example of concentric circles, we also saw that the discrete solutions overshoot when they reach the t -axis. The reason is *concavity*. Suppose y is a solution above the t -axis, then $y > 0$ and $y' < 0$. Then, we can find the sign of the second derivative of y :

$$y'' = \frac{d}{dt}f(t, y) = \frac{d}{dt}\left(-\frac{t}{y}\right) = -\frac{1 \cdot y - t \cdot y'}{y^2} = \frac{t \cdot y' - y}{y^2} < 0.$$

It's negative! It's positive below the t -axis. Therefore, the solutions are concave down in the former and concave up in the latter case. Then, the linear approximations that we use for Euler's method overestimate the solutions in the former case and underestimate in the latter.



The benefit is that we are cutting the tunnel in half.

Corollary 1.10.12: One-sided Error Bound

Under the conditons of the last theorem, we have the following:

- When $f' > 0$, we have for each t in $[t_0, t_0 + H]$:

$$y_n \leq y(t_0 + nh) \leq y_n + \frac{1}{2}KHh.$$

- When $f' < 0$, we have for each t in $[t_0, t_0 + H]$:

$$y_n - \frac{1}{2}KHh \leq y(t_0 + nh) \leq y_n.$$

Exercise 1.10.13

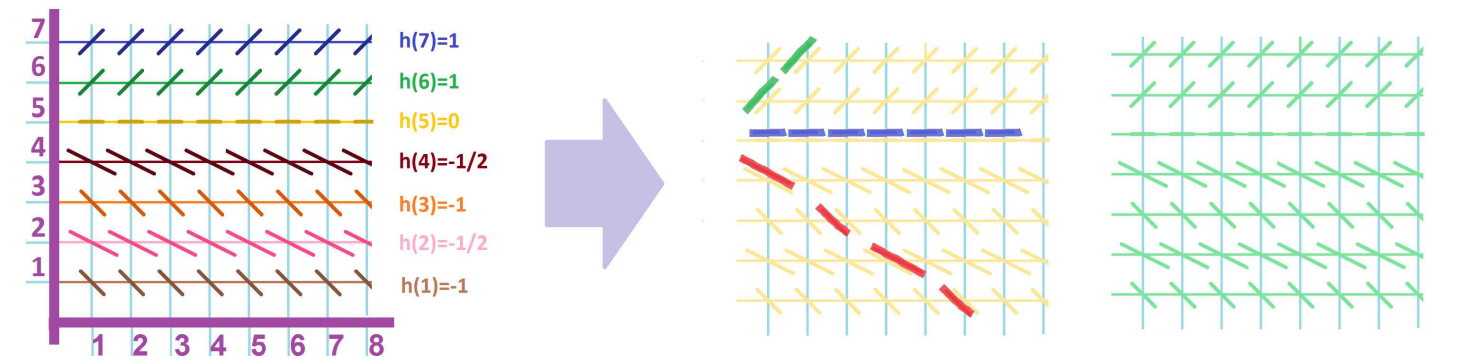
Apply the corollary to the example of concentric circles in the beginning of the section.

1.11. Qualitative analysis of ODEs

Unfortunately, Euler’s method depends on the value of h and, from simple experimentation, one cannot know whether h is small enough. It’s just an approximation! The method can also fail near the boundary of the domain.

With these limitation of the *quantitative* methods, *qualitative* analysis provides fully accurate if broad descriptions of the solutions. The method amounts to gathering information about the solutions without solving the ODE – either analytically or numerically.

As we saw early in this chapter, some ODEs $y' = f(y)$ have the right-hand side function independent of t as in $f(y) = y^2$, etc. This means that the field of slopes on the xy -plane under horizontal shift will land on itself:



Then so will its solution set! The following will be very important.

Theorem 1.11.1: Time Independent ODEs

Suppose the right-hand side of an ODE $y' = f(y)$ is independent of t . Then, if $y = y(t)$ is a solution of the ODE then so is $y = y(t + s)$ for any real s .

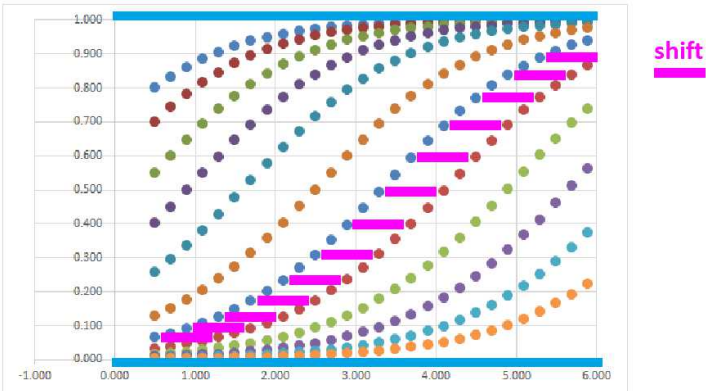
Exercise 1.11.2

Prove the theorem.

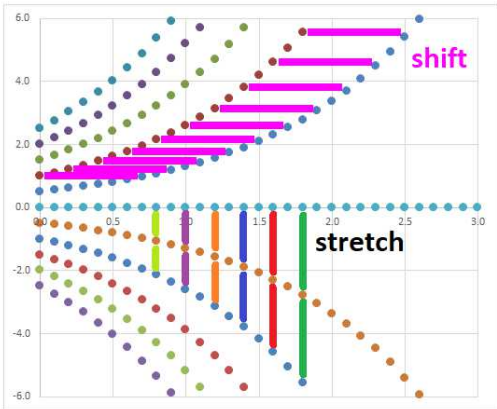
Example 1.11.3: logistic equation

Consider the logistic equation:

y\ t	t																				f
	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1.00
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.95	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
0.90	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
0.85	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
0.80	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
0.75	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
0.70	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1
0.65	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1
0.60	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2
0.55	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2
0.50	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3	1.3
0.45	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2
0.40	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2
0.35	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1
0.30	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1
0.25	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
0.20	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
0.15	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
0.10	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
0.05	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
0.00	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0



Consider also the population model:



Not only it is preserved under horizontal shifts, but also under vertical stretches!

The main tools come from Volume 2 (the Monotonicity Theorem, [Chapter 2DC-5](#)). First, a differentiable function y defined on an open interval is increasing when its derivative is positive and decreasing when it's negative. In other words, we have:

$$\begin{aligned} y' > 0 &\implies y \nearrow \\ y' < 0 &\implies y \searrow \end{aligned}$$

Theorem 1.11.4: Monotonicity of Solutions

If y is a solution of the ODE $y' = f(t, y)$ within an open set in the ty -plane, we have on this set:

$$\begin{aligned} f(t, y) > 0 &\implies y \text{ is increasing.} \\ f(t, y) < 0 &\implies y \text{ is decreasing.} \end{aligned}$$

The second tool is also elementary.

Theorem 1.11.5: Stationary Solutions

If $f(t, y_0) = 0$ for all t within an open interval and some y_0 , then $y(t) = y_0$ is a (constant) solution of the ODE $y' = f(t, y)$.

Example 1.11.6: trigonometric

Consider:

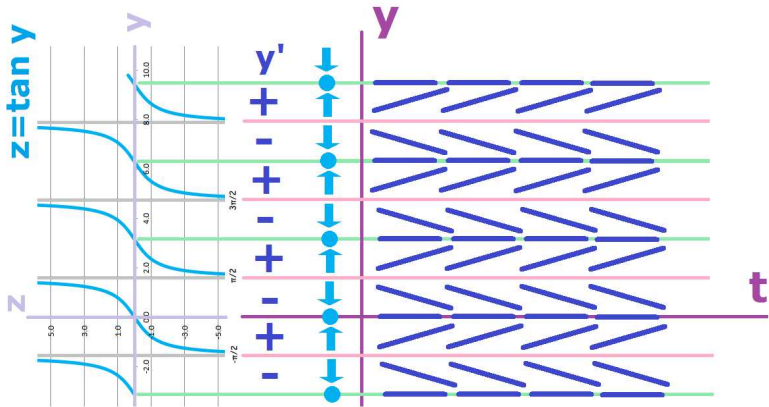
$$y' = -\tan(y).$$

First, the right-hand side, the tangent, is undefined at $y = \pi/2 + k\pi$ for all integer k . Therefore, these

horizontal lines on the ty -plane will cut the domain into strips so that every solution will stay inside one of them. Furthermore, the sign of the tangent changes at those values of y as well as at $y = k\pi$, in the middle of each strip. For example, this is what we conclude about the strip $(-\pi/2, \pi/2)$:

- For $-\pi/2 < y < 0$, we have $y' = -\tan y > 0$ and, therefore, $y \nearrow$.
- For $y = 0$, we have $y' = -\tan y = 0$ and, therefore, y is a constant solution.
- For $0 < y < \pi/2$, we have $y' = -\tan y < 0$ and, therefore, $y \searrow$.

The results of the analysis are summarized below with the monotonicity of the solution matched with the sign of the tangent:



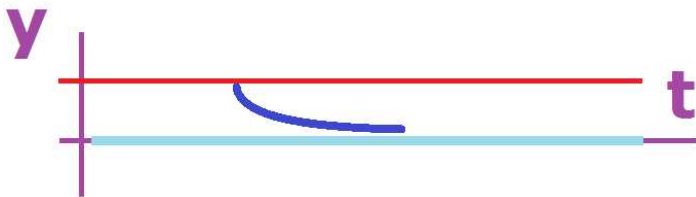
So, every solution y is decreasing (or increasing respectively) throughout its domain. There are more subtle conclusions. According to a theorem, the existence is satisfied within the strip. Therefore, the solutions can be extended further and further. According to another theorem, the uniqueness is satisfied within the strip. Therefore, then the solutions can't cross the line $y = 0$. We conclude:

- Every solution is approaching the middle line of the strip, $y = 0$, as $t \rightarrow +\infty$, and this line is a horizontal asymptote of the solution, i.e., $y \rightarrow 0^+$.

But how does it approach the no-solution lines $y = \pi/2 + k\pi$? As the tangent is approaching infinity, so is the slope of the solution. We conclude:

- Every solution is approaching the edge lines of the strip, $y = \pm\pi/2$, and it becomes more and more vertical, i.e., $y' \rightarrow \pm\infty$.

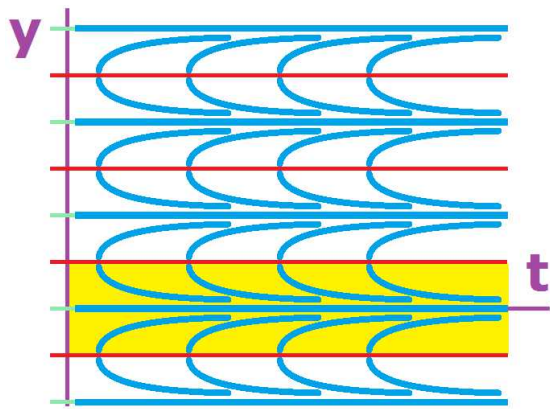
This is what the solution set might look like when drawn by hand:



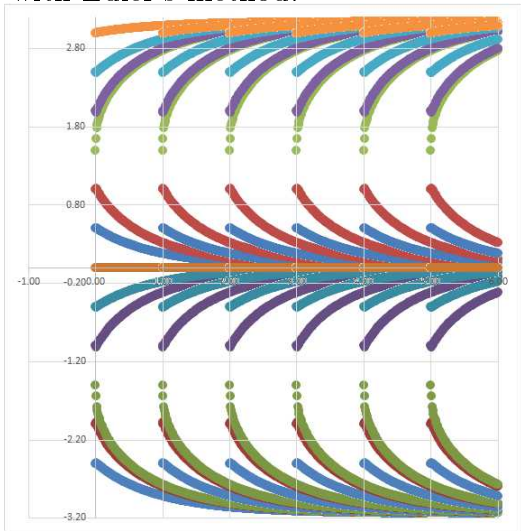
Now we utilize some facts about the function f .

1. We use that fact that f is independent of t to produce (according to the theorem above) more solutions by a horizontal shifts.
2. We use the fact that f is odd to produce more solutions by a vertical flip.
3. We use the fact that f is periodic to produce more solutions by vertical shifts.

This is the solution set:



The conclusions are confirmed with Euler’s method:



We notice now that the solutions are running away from the edge of the domain and toward the constant solution. We also notice that the solutions are repeated under a horizontal shift.

Warning!

We confirm our qualitative analysis with Euler’s method... and vice versa!

Exercise 1.11.7

In the last example, is $y = \pi/2$ also a horizontal asymptote? Confirm your conclusion with Euler’s method. What is y' near this line?

Notice that both above and below Euler’s method is run for various values of $t_0 = 0, 1, 2, 3 \dots$ in order to avoid empty spaces created by the asymptotic behavior of the solutions.

Example 1.11.8: more complex

Consider next:

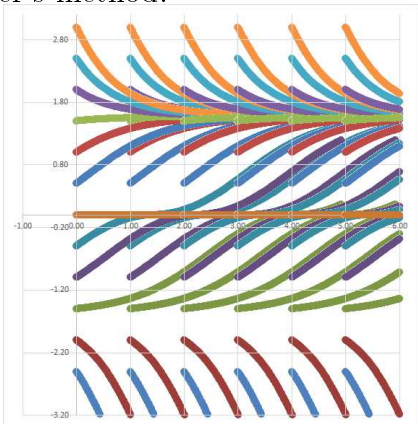
$$y' = \cos y \cdot \sqrt{|y|}.$$

We start with the “sign analysis” of the right-hand side, $f(y) = \cos y \cdot \sqrt{|y|}$, which is fully determined

by $\cos y$ unless $y = 0$:

y	$f(y)$	$y = y(t)$	
$3\pi/2$	0	$y(t) = 3\pi/2$, constant	$\rightarrow\rightarrow\rightarrow\rightarrow$
	−	$y \searrow$	$\searrow\searrow\searrow\searrow$
$\pi/2$	0	$y(t) = \pi/2$, constant	$\rightarrow\rightarrow\rightarrow\rightarrow$
	+	$y \nearrow$	$\nearrow\nearrow\nearrow\nearrow$
0	0	$y(t) = 0$, constant	$\rightarrow\rightarrow\rightarrow\rightarrow$
	+	$y \nearrow$	$\nearrow\nearrow\nearrow\nearrow$
$-\pi/2$	0	$y(t) = -\pi/2$, constant	$\rightarrow\rightarrow\rightarrow\rightarrow$
	−	$y \searrow$	$\searrow\searrow\searrow\searrow$
$-3\pi/2$	0	$y(t) = -3\pi/2$, constant	$\rightarrow\rightarrow\rightarrow\rightarrow$

The results are confirmed with Euler’s method:



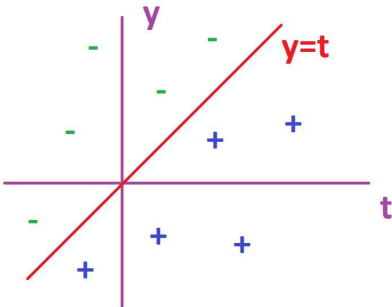
The plotted Euler solutions cross the path of $y = 0$ indicating *non-uniqueness*. Considering non-differentiability of $f(y)$ at $y = 0$, this is a possibility. Elsewhere, however, the function is differentiable and the uniqueness holds. The behavior is asymptotic just as in the last example.

Example 1.11.9: non-homogeneous

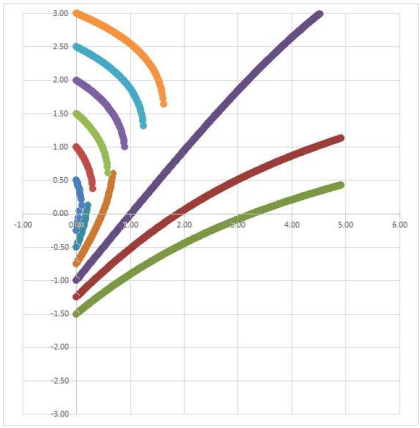
The next one is dependent on t :

$$y' = \frac{1}{t - y}.$$

The right-hand side is undefined whenever $t - y = 0$. Therefore, no solution crosses the diagonal line $y = t$. Furthermore, the solutions above it are decreasing and the ones below it are increasing as follows:



Check:



Example 1.11.10: trig non-homogeneous

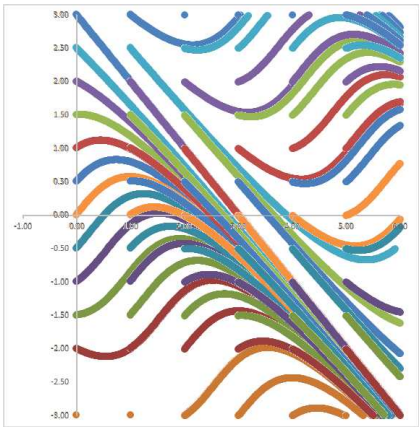
The next one is defined on the whole plane:

$$y' = \cos(t + y) .$$

The sign is changing whenever $t + y = \pi/2 + k\pi$ for all integer k . Thus the solution change their monotonicity whenever they reach one of the lines

$$y = -t + \pi/2 + k\pi$$

of slope -1 . One of these lines however is special and we wouldn't think of it without doing Euler's method first:



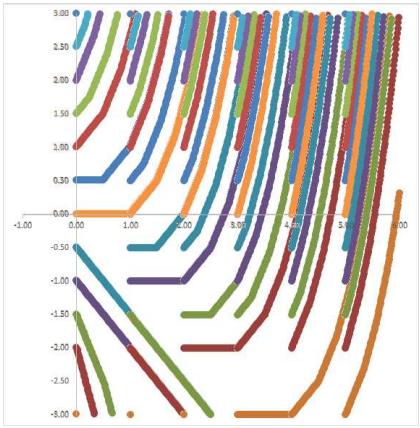
Indeed, $y = \pi - t$ is a solution! The rest of them are waves moving diagonally.

Example 1.11.11: discontinuous RHS

The next one is discontinuous:

$$y' = [t + y] .$$

Since the function changes abruptly, so does the derivative (the slope) of each solution. For Euler's method we use the `FLOOR` function:

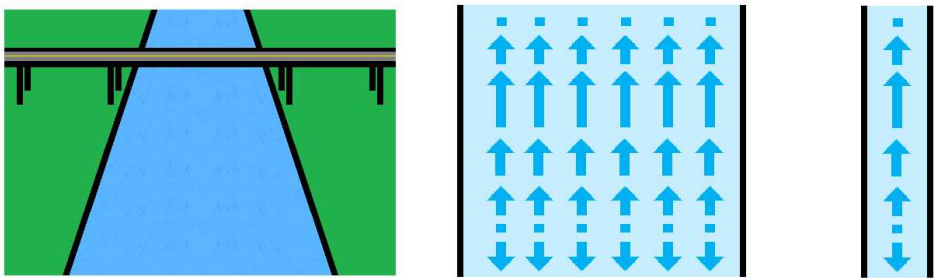


The corners indicate that these are weak solutions.

Suppose the ODE is *time-independent*,

$$y' = f(y) .$$

Then it is thought of a liquid flow: in a canal or a pipe.

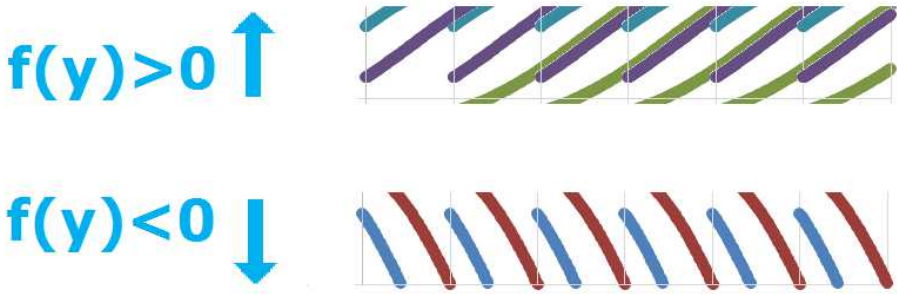


Especially in the latter case, the right-hand side is recognized as a *one-dimensional vector field*.

With such an equation, the qualitative analysis is much simpler. In fact, one of the ODEs above, $f(y) = \cos y \cdot \sqrt{|y|}$, exhibits all possible patterns of *local* behavior.

We concentrate on what is going on in the vicinity of a given location $y = a$.

The first main possibility is $f(a) \neq 0$. Then, from the continuity of f , we conclude that $f(y) > 0$ or $f(y) < 0$ in some interval I that contains a . Then, the solutions located within the band $(-\infty, +\infty) \times I$ are either all increasing or all decreasing respectively:



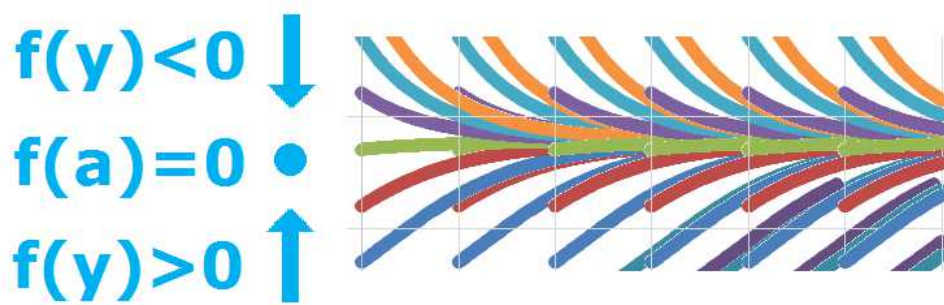
The behavior is “generic”.

More interesting behaviors are seen around a zero of f :

► $f(a) = 0 \implies y = a$ is a stationary solution (an equilibrium).

Then the pattern in the vicinity of the point, i.e., an open interval I , depends on whether this is a maximum of f , a minimum, or neither.

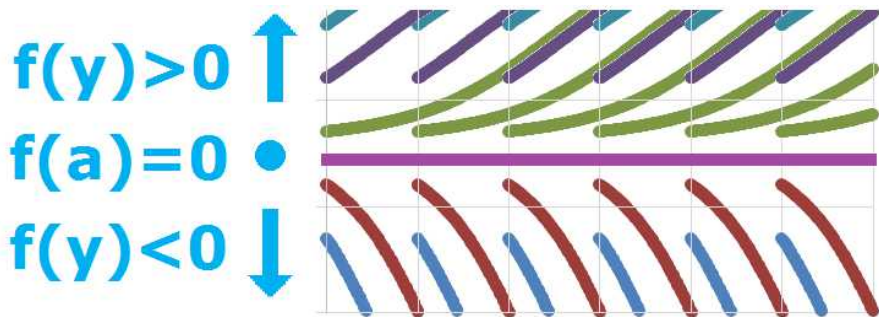
First, if f changes its sign from positive to negative, we have a *stable equilibrium*, or a “sink”:



Indeed, we have all solutions within the band $(-\infty, +\infty) \times I$, under the uniqueness condition, asymptotically approach $y = a$:

$$y \rightarrow a \text{ as } x \rightarrow +\infty .$$

Second, if f changes its sign from negative to positive, we have an *unstable equilibrium*, or a “source”:



It means that there are no solutions within the band $(-\infty, +\infty) \times I$, under the uniqueness condition, that asymptotically approach $y = a$ as $x \rightarrow +\infty$. On the other hand, we have

$$y \rightarrow a \text{ as } x \rightarrow -\infty .$$

Third, if f does not change its sign at $y = a$, we have an *semi-stable equilibrium*, or a “pass”:

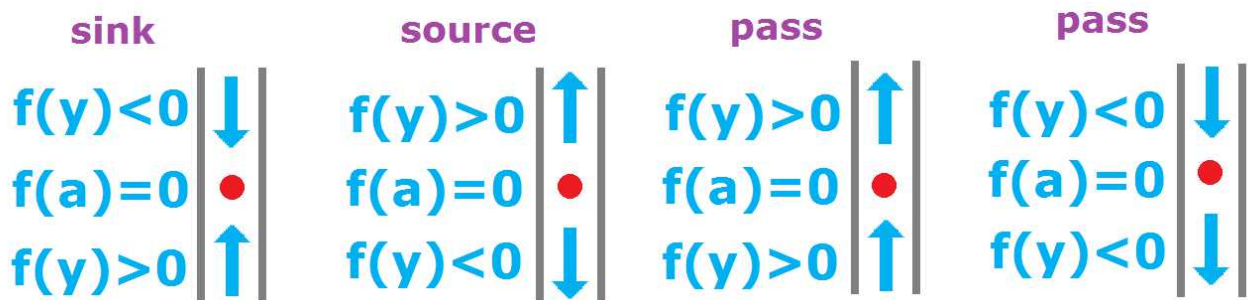


It means that the solutions within one of the two sides of the line $y = a$ within the band $(-\infty, +\infty) \times I$, under the uniqueness condition, asymptotically approach this line and the ones on the other side do not.

The last possibility is $f(y) = 0$ on the whole I . Then all the solutions in the band are stationary. This is a *degenerate equilibrium*.

This *classification of equilibria* is easy to understand in terms of the flow represented by the ODE:

- sink: flow in only (stable equilibrium);
- source: flow out only (unstable equilibrium);
- pass: flow, stop, flow (semi-stable equilibrium).



Warning!

In spite of the names, a particle placed at a source will never leave it and a particle placed near a sink will never reach it, under the uniqueness condition.

There are only these three main possibilities in the 1-dimensional case. We will see in Chapter 3 a wide variety of behaviors around an equilibrium when the space of location is 2-dimensional, a plane.

1.12. Linearization of ODEs

Euler’s method approximates solutions of an ODE by linearizing them, one location at a time. We can also approximate the solutions of the ODEs by linearizing, locally, its right-hand side.

The main idea is the following. As we are often unable to solve a differential equation we encounter, let’s try to replace it with a simpler one that can be solved. How accurate do we need the approximation to be? At the very least, the new ODE should have the same *qualitative behavior* – increasing, decreasing, and stationary solutions – in the vicinity of the chosen point.

We will focus on the time-independent ODE with a continuous right hand-side,

$$y' = f(y) \, ,$$

in the vicinity of a chosen point $y = a$. We will rely on the results stated previously.

Before we get to the linear, let’s try the *best constant approximation*.

What is the best constant approximation at $x = a$ of a continuous function $y = f(x)$? It’s $y = f(a)$!

Example 1.12.1: simplest approximation

Solve approximately:

$$y' = y^2 \text{ around } y = 1 \, .$$

We evaluate the function at that point:

$$y^2 \Big|_{y=1} = 1 \, .$$

Then, we replace our function with this constant function $g(y) = 1$. We have a new, and very simple, ODE:

$$y' = 1 \, .$$

Its solution is:

$$y = x + C \, .$$

It’s a poor substitute. However, the qualitative behavior is the same!

However, let’s solve the same ODE around another point:

$$y' = y^2 \text{ around } y = 0 \, .$$

We evaluate the function at that point:

$$y^2 \Big|_{y=0} = 0 \, .$$

Then, we replace our function with this constant function $g(y) = 0$. We have a new, and very simple, ODE:

$$y' = 0 \, .$$

Its solution is constant:
$$y = C.$$
It's an unacceptably bad substitute because the qualitative behavior is the different!

In general, for a chosen point $y = a$, we replace $z = f(y)$ with $z = f(a)$:

► Instead of $y' = f(y)$, we solve $y' = f(a)$.

In other words, the value $f(a)$ of f at a is used the right-hand side of the new ODE:

$$y' = f(a),$$

for all y in some open interval I that contains a . Its solutions are these *linear* functions of t :

$$y = f(a)(t + C),$$

for each real C .

Does their behavior match the original? There are two main cases:

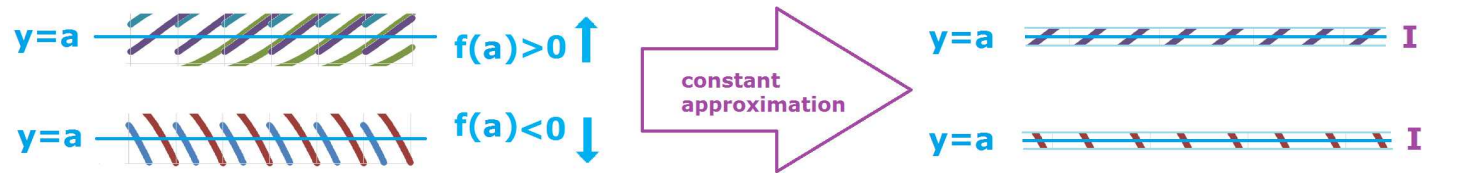
- 1. $f(a) \neq 0$
- 2. $f(a) = 0$

This is what we conclude about *all* solutions of the simplified ODE.

In the former case, we have:

$$f(a) > 0 \implies y \nearrow \text{ and } f(a) < 0 \implies y \searrow,$$

for all solutions in the band $(-\infty, +\infty) \times I$. They are curved on the left and straight on the right:



Indeed, to compare to the solutions of the original ODE, we can just apply the Monotonicity Theorem from calculus. It's a match!

Below is the summary of this case.

Theorem 1.12.2: Constant Approximation of ODEs

Suppose $z = f(y)$ is continuous at $y = a$. Suppose

$$f(a) \neq 0.$$

Then the qualitative behavior of the solutions of the ODE $y' = f(y)$ match those of its constant substitute $y' = f(a)$, as follows: there is such an $\varepsilon > 0$ that each solution y of $y' = f(y)$ the graph of which lies within the band $(-\infty, +\infty) \times (a - \varepsilon, a + \varepsilon)$ satisfies:

$$\begin{aligned} f(a) > 0 &\implies y \text{ is increasing.} \\ f(a) < 0 &\implies y \text{ is decreasing.} \end{aligned}$$

Proof.

The continuity of f implies that there is $\varepsilon > 0$ small enough that for each y within $(a - \varepsilon, a + \varepsilon)$, we

have

$$\begin{aligned} f(a) > 0 &\implies f(y) > 0 \\ f(a) < 0 &\implies f(y) < 0 \end{aligned}$$

In the latter case, we have:

$$f(a) = 0 \implies y \text{ is constant .}$$



Then all the solutions in the band are stationary. But the original ODE might have a semi-stable equilibrium, a mismatch!

Conclusion:

- The constant approximation of $y = f(y)$ at $x = a$ fails when $f(a) = 0$.

To improve our chances, we try is the *best linear approximation* of f .

We concentrate on the case, of course, that has not been, qualitatively, solved by the best constant approximation, i.e.,

$$f(a) = 0 .$$

Example 1.12.3: around stationary point

Let’s solve approximately:

$$y' = y^2 + y \text{ around } y = 1 .$$

We evaluate the function at that point:

$$y^2 + y \Big|_{y=1} = 2 .$$

Then, we replace our function with this constant function $g(y) = 2$. We have a new, and very simple, ODE:

$$y' = 2 .$$

Its solution is:

$$y = 2x + C .$$

The qualitative behavior is the same according to the theorem. The constant approximation is good enough!

Let’s solve approximately elsewhere:

$$y' = y^2 + y \text{ around } y = 0 .$$

We evaluate the function at that point:

$$y^2 + y \Big|_{y=0} = 0 .$$

Then, we replace our function with this constant function $g(y) = 0$. We have a new, and very simple, ODE:

$$y' = 0 .$$

Its solution is:

$$y = C .$$

The qualitative behavior is not the same. The constant approximation is good enough!

Now linear approximation. We evaluate the derivative:

$$(y^2 + y)' \Big|_{y=0} = 2y + 1 \Big|_{y=0} = 1.$$

Then, we replace our function with this linear function $g(y) = y$. We have a new, and very simple, ODE:

$$y' = y.$$

Its solution is:

$$y = Ce^x.$$

The qualitative behavior is the same. The linear approximation is good enough!

Now the general case. For simplicity, the process of linearization starts with moving the point of interest, currently $y = a$, to 0. Such a substitution is demonstrated earlier in the chapter. This is the new ODE:

$$y' = f(y) \text{ with } f(0) = 0.$$

The linearized ODE is

$$y' = f'(0)y,$$

for all y in some open interval I that contains 0. Its solutions are these *exponential* functions of t :

$$y = Ce^{f'(0)t},$$

for each real C .

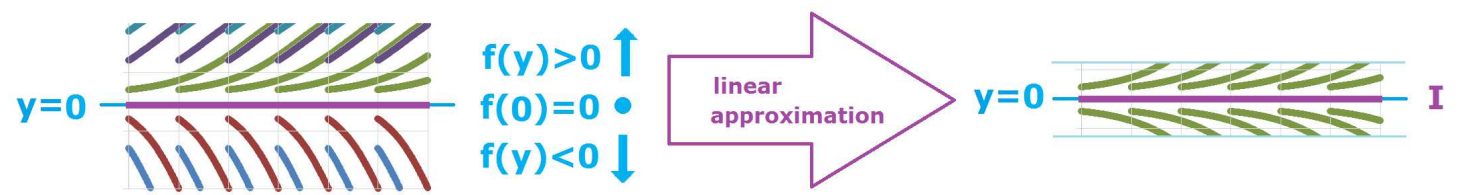
Does their behavior match the original? There are two main cases:

- 1. $f'(0) \neq 0$
- 2. $f'(0) = 0$

This is what we conclude about *all* solutions of the linearized ODE. First we have:

$$f'(0) < 0 \implies f \searrow \implies f(y) > 0 \text{ for } y < 0 \text{ and } f(y) < 0 \text{ for } y > 0,$$

for all solutions in the band $(-\infty, +\infty) \times I$. Then, f changes its sign from positive to negative and we have an unstable equilibrium:



Unlike the constant approximation, the linearization has produced a match!

Second, we have within the band:

$$f'(0) > 0 \implies f \nearrow \implies f(y) < 0 \text{ for } y < 0 \text{ and } f(y) > 0 \text{ for } y > 0.$$

Then, f changes its sign from negative to positive and we have a stable equilibrium. Once again, a match!

Below is the summary of this case.

Theorem 1.12.4: Linear Approximation of ODEs

Suppose $z = f(y)$ is differentiable at $y = 0$. Suppose

$$f(0) = 0 \text{ and } f'(0) \neq 0.$$

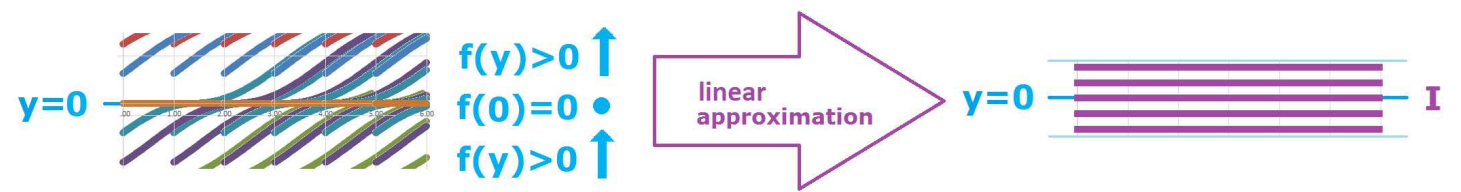
Then the qualitative behavior of the solutions of the ODE $y' = f(y)$ match those of its linear substitute $y' = f'(0)y$, as follows: there is such an $\varepsilon > 0$ that each solution y of $y' = f(y)$ the graph of which lies within the band $(-\infty, +\infty) \times (\varepsilon, \varepsilon)$ satisfies:

$f'(0) > 0 \implies y \text{ is decreasing when } y < 0 \text{ and } y \text{ is increasing when } y > 0.$ $f'(0) < 0 \implies y \text{ is increasing when } y < 0 \text{ and } y \text{ is decreasing when } y > 0.$

Exercise 1.12.5

Linearize the ODE $y' = ye^y$ at $y = 0$.

What about the case $f'(0) = 0$? We are in a similar place to the one for the constant approximation. Here, the linearization gives us the ODE $y' = 0$ with stationary solutions only. In the meantime, f might have a semi-stable equilibrium:



A mismatch!
Conclusion:

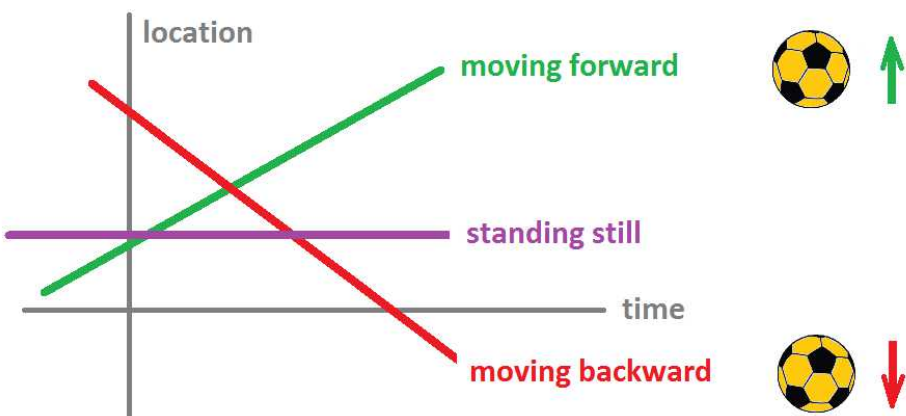
- The best linear approximation of $y = f(y)$ at $x = 0$ fails when $f(0) = 0$ and $f'(0) = 0$.

The answer is the best *quadratic* approximation, i.e., the second Taylor polynomial T_2 of f around $y = 0$, provided f is twice differentiable. However, there may still be exceptions that will call for using the *cubic* Taylor polynomial T_3 of f . And so on. We will need all the Taylor polynomials, i.e., the *Taylor series*. The idea is developed in [Chapter 2](#).

1.13. Motion under forces: the acceleration

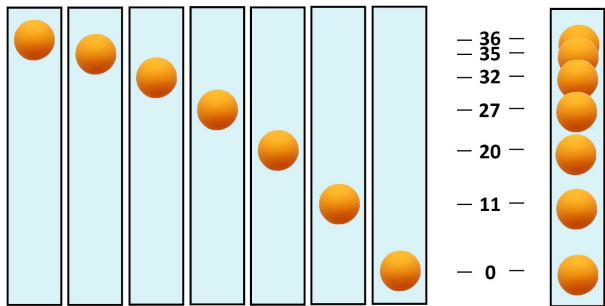
Example 1.13.1: moving ball

We know that a *ball rolling* on a horizontal plane will have a constant velocity:

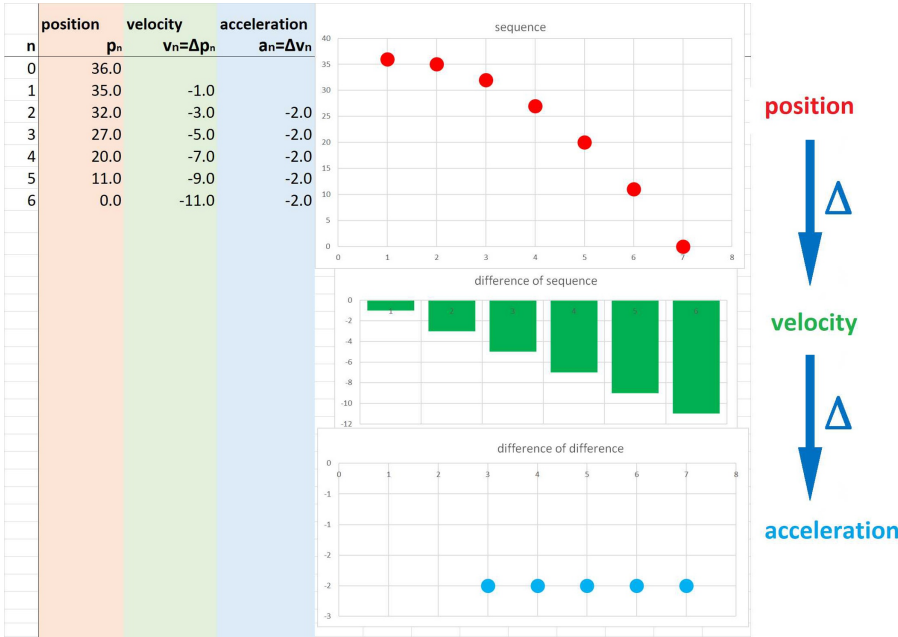


What if the ball is now thrown *up in the air*? The dynamics is very different. In the former, as there is no force changing the velocity, the latter remains constant. In the latter, the velocity is constantly changed by the gravity.

Imagine that we have this experimental data of the heights of a ping-pong ball falling down recorded about every .1 second measured in inches:



We use a spreadsheet to plot the *location* sequence, p_n (red). We then compute the difference of p_n , i.e., the *velocity*, v_n (green):



It looks like a straight line. But this time, we take one more step: We compute the difference of the velocity sequence. It is the *acceleration*, a_n (blue). It appears constant! There might be a law of nature here.

Let’s accept the premise we’ve put forward:

► *The acceleration of free fall is constant.*

Then we can try to predict the behavior of an object thrown in the air – from any initial height and with any initial velocity. The direction of our computation is opposite to that of the last example:

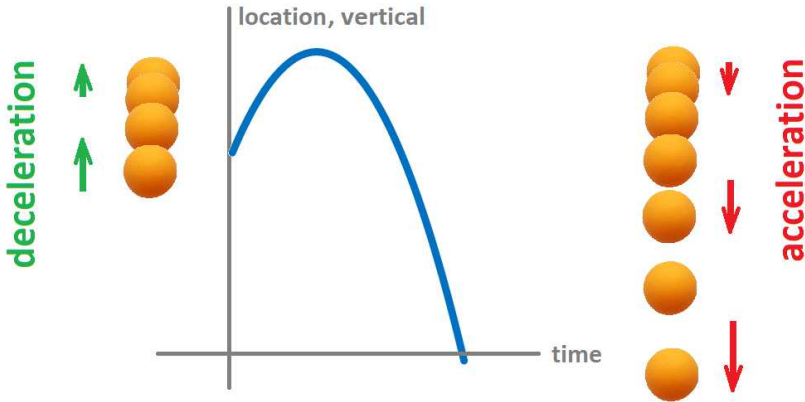
► We use our knowledge of the acceleration to derive the velocity, and then derive the position of the object in time.

While we used *differences* in the last example, we use *sums* (Chapter 1PC-1) now.

We plot these positions against time:



The graph “looks like” a parabola:



We use these difference quotients formulas to find the velocity from the position and then the acceleration from the velocity:

$$v_n = \frac{\Delta p}{\Delta t} = \frac{p_{n+1} - p_n}{h} \quad \text{and} \quad a_n = \frac{\Delta v}{\Delta t} = \frac{v_{n+1} - v_n}{h},$$

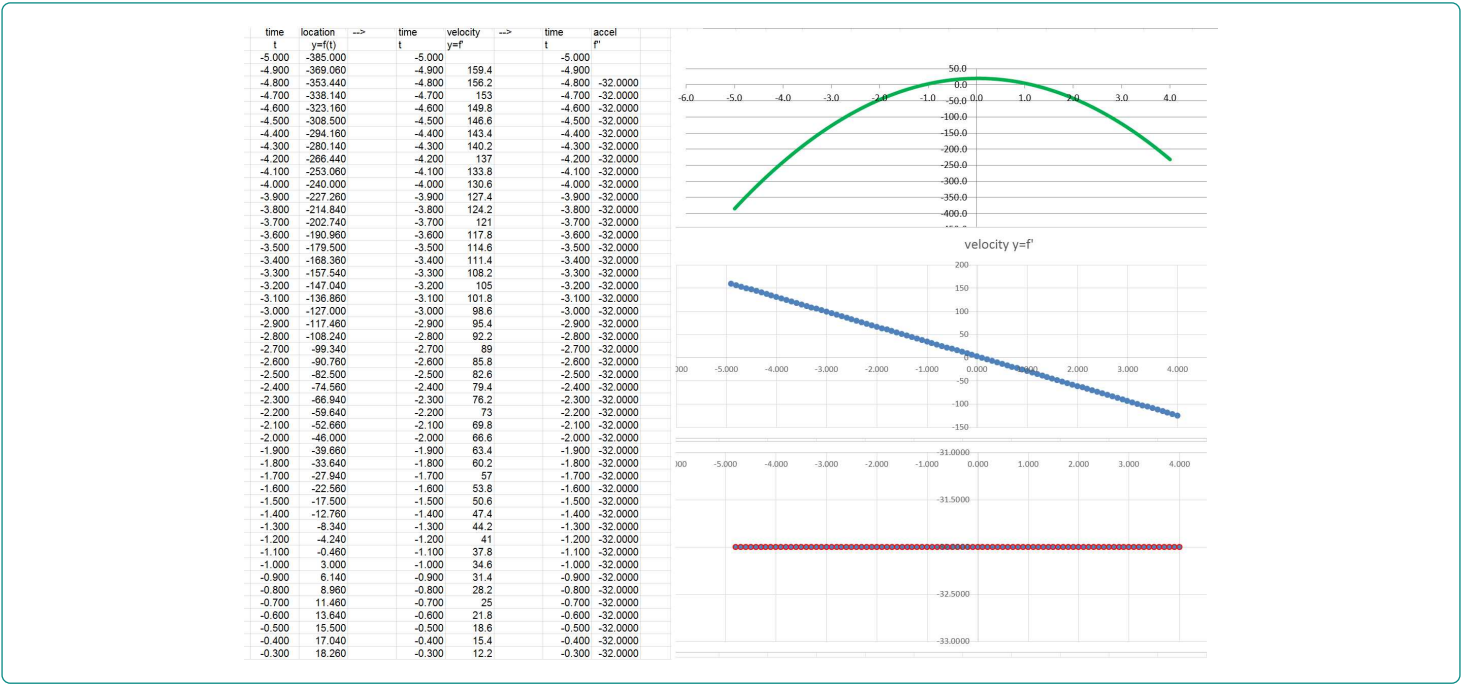
where h is the increment of time. The dependence of the velocity on the position and the acceleration on the velocity is, of course, identical.

Example 1.13.2: spreadsheet

The following formula is used again:

`= (RC[-1] - R[-1]C[-1]) / R2C1`

These are the results:



Now in reverse! For simulation, the derivation goes in the opposite direction:

- the velocity from the acceleration and then
- the location from the acceleration.

The formulas are solved for p_{n+1} and v_{n+1} respectively:

$$v_n = \frac{p_{n+1} - p_n}{h} \implies v_{n+1} = v_n + ha_n$$
$$a_n = \frac{v_{n+1} - v_n}{h} \implies p_{n+1} = p_n + hv_n$$

The dependence of the velocity on the acceleration and the position on the velocity is, of course, identical.

Summary:

	vertical			vertical
position	p_n given	\implies	acceleration	a_n given
velocity	$v_n = \frac{p_{n+1} - p_n}{h}$		velocity	$v_{n+1} = v_n + ha_n$
acceleration	$a_n = \frac{v_{n+1} - v_n}{h}$		position	$p_{n+1} = p_n + hv_n$

Example 1.13.3: moving ball

Let’s consider a specific problem.
► **PROBLEM:** From a 100 feet building, a ball is thrown up at 50 feet per second so that it falls on the ground. How high will the ball go?
We use the same spreadsheet formula for the velocity and position:

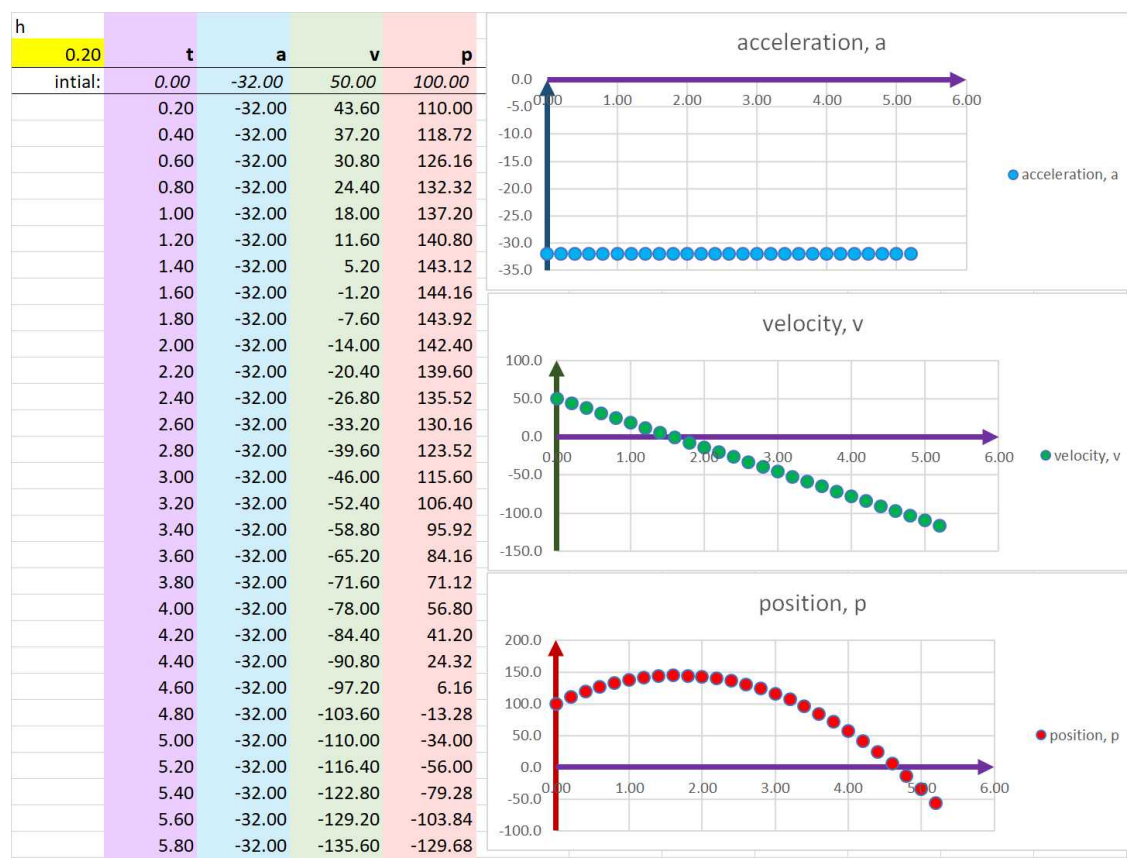
=R[-1]C+R[-1]C[-1]*R2C1

Now the specific case of *free fall*, there is just one force, the gravity, and the vertical acceleration is known to be $a = -g$, where g is the gravitational constant:
$$g = 32 \text{ ft/sec}^2.$$

Next, we acquire the initial conditions:

- The initial location is given by: $p_0 = 100$.
- The initial velocity is given by: $v_0 = 50$.

We use the formulas to evaluate the location every $h = .20$ second. This is what the graphs look like:



With the spreadsheet, we can ask and answer a variety of other questions about such motion (how hard it hits the ground, etc.). However, we can only do one example at a time! The conclusions we draw are also specific to these initial conditions.

We now consider the *continuous* case, i.e., we take the limit of everything above:

$$h = \Delta x \rightarrow 0.$$

Instead of sequences, we have this time these *functions of time*:

- p is the height, the vertical location.
- $v = p'$ is the vertical velocity.
- $a = v'$ is the vertical acceleration.

Now the specific case of free fall:

$$a = -g.$$

We know that:

1. The derivative of a quadratic polynomial is linear.
2. The derivative of a linear polynomial is constant.

Conversely:

1. The only function the derivative of which is linear is a quadratic polynomial.
2. The only function the derivative of which is constant is a linear polynomial.

We conclude that

► $p = p(t)$ is quadratic.

In other words, we have:

$$p(t) = ax^2 + bx + c.$$

What makes these specific are the *initial conditions*:

- p_0 is the initial height, $p_0 = p(0)$.
- v_0 is the initial vertical component of velocity, $v(0) = \frac{dp}{dt}\Big|_{t=0}$.

Therefore, we have:

$$p(t) = p_0 + v_0 t - \frac{1}{2}gt^2$$

Example 1.13.4: moving ball

In the problem of ours, we have:

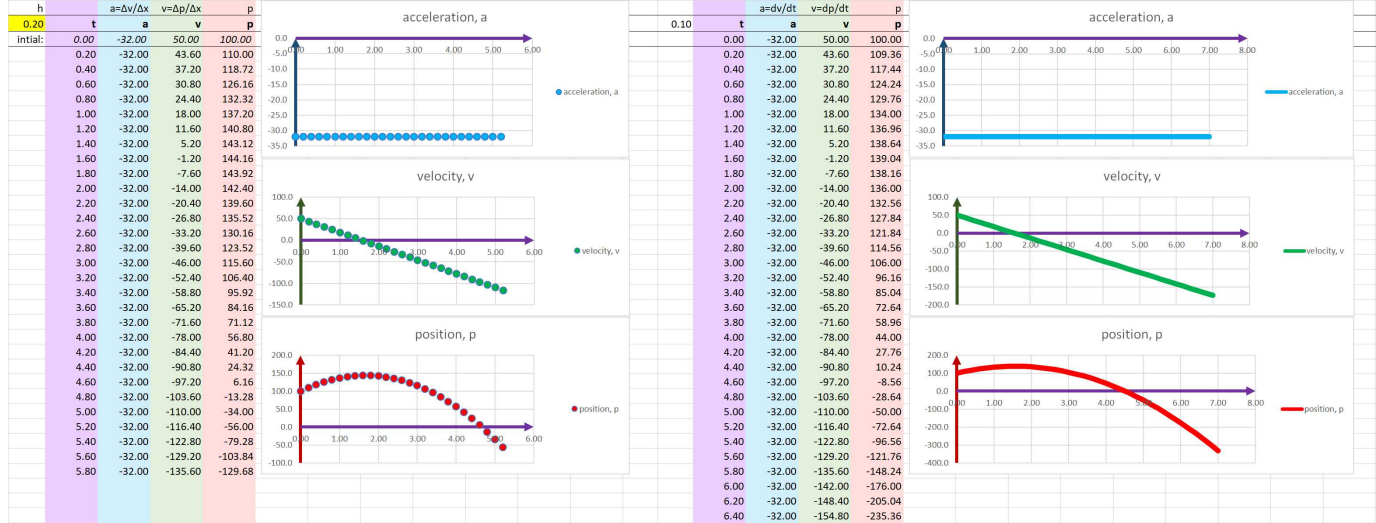
$p_0 = 100, v_0 = 50.$

Our equations become:

$p = 100 + 50t - 16t^2.$

In contrast to the discrete case, the formula for the position isn't recursive but direct and explicit!

Before we utilize the explicit algebraic representation, we visualize the results by plotting the graph of this function next to the one obtained recursively:



Exercise 1.13.5

What happened to \pm ?

Exercise 1.13.6

How high does the projectile go in the above example?

Exercise 1.13.7

Using the above example, how long will it take for the projectile to reach the ground if fired *down*?

Exercise 1.13.8

Use the above model to determine how long it will take for an object to reach the ground if it is dropped. Make up your own questions about the situation and answer them. Repeat.

Example 1.13.9: acceleration that depends on time

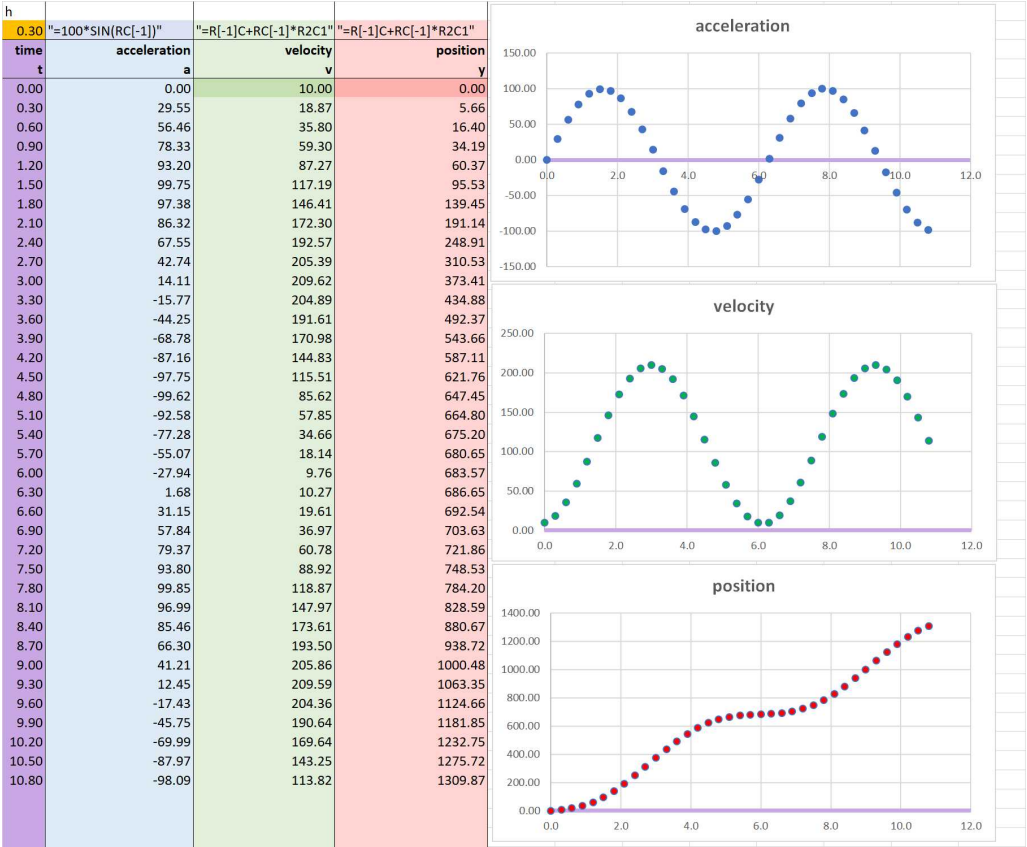
What your algebraic analysis is incapable of doing (for now) is to handle the case of time-dependent acceleration. For example, imagine that the Earth moves in and out – periodically – so close to the Sun that the gravity of the latter is stronger than that of the former. Then the acceleration will be

also changing periodically. What will be the motion of the same ball?

The spreadsheet works with no change! We just insert a periodic formula for the acceleration:

=100*SIN(RC[-1])

Then the rest is taken care of:



Exercise 1.13.10

Devise such initial conditions for the last example that the object oscillates between the Earth and the Sun.

1.14. Discrete models: how to set up ODEs of second order

We have already seen the simplest (first order) ODEs for motion: if we know the *velocity* of a moving object, we can predict its dynamics.

Suppose instead what we know is the *forces* affecting the object and, therefore, its acceleration. How can we predict its dynamics?

Let's review the *discrete model*. We start with the following four quantities that come from the setup of the motion:

- the initial time t_0
- the initial velocity v_0
- the initial location p_0
- the current acceleration a_1

We choose:

$$h = \Delta t.$$

As we progress in time and space, new numbers are placed in the next row of our spreadsheet:

$$t_n, \; a_n, \; v_n, \; p_n, \; n = 1, 2, 3, \dots$$

with the following recursive formulas:

- $t_{n+1} = t_n + h$
- $v_{n+1} = v_n + a_n \cdot h$
- $p_{n+1} = p_n + v_n \cdot h$

The result is a growing table of values:

	iteration n	time t_n	acceleration a_n	velocity v_n	location p_n
initial:	0	3.5	—	33	22
	1	3.6	66	38.5	25.3

	1000	103.5	666	4	336

Where does the current acceleration come from? Our main interest is the case when the acceleration depends on the last location, e.g., $a_{n+1} = 1/p_n^2$, such as when the gravity depends on the distance to the planet or the force of the spring depends on the distance of its end from the equilibrium.

Example 1.14.1: no forces

Recall some examples. A rolling ball is unaffected by horizontal forces and the recursive formulas for the horizontal motion simplify as follows:

- The velocity $v_{n+1} = v_n + a_n \cdot h = v_n = v_0$ is constant.
- The position $p_{n+1} = p_n + v_n \cdot h = p_n + v_0 \cdot h$ grows at equal increments.

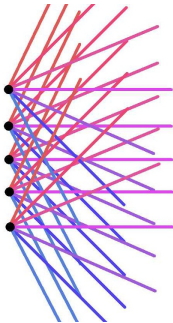
In other words, the position depends linearly on the time. As $\Delta x \rightarrow 0$, we have an equation:

$$y'' = 0.$$

The solution set looks is:

$$y = p_0 + v_0 t,$$

taken over all possible initial conditions. It can be illustrated as follows:



All linear functions are here! This is a *two*-parameter family: the y -intercepts and the slopes.

Example 1.14.2: free fall

A falling ball is unaffected by horizontal forces and the vertical force is constant: $a_n = a$ for all n . The first of the two recursive formulas for the vertical motion simplifies as follows:

- The velocity $v_{n+1} = v_n + a_n \cdot h = v_n + a \cdot h$ grows at equal increments.
- The position $p_{n+1} = p_n + v_n \cdot h$ grows at linearly increasing increments.

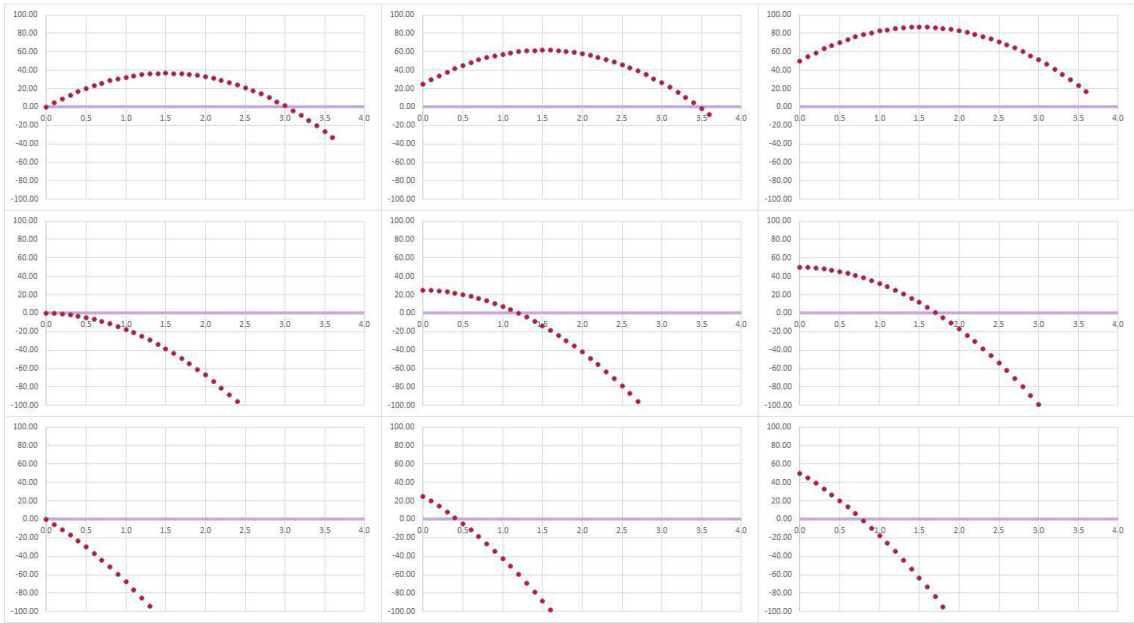
Finally, if we think of y as a sampled twice differentiable function, we have a single ODE:

$$y'' = -g.$$

The solution set looks is:

$$y = p_0 + v_0t - \frac{1}{2}gt^2,$$

taken over all possible initial conditions. This is a sketch of the solution set:



This is also a *two*-parameter family: the y -intercepts and the slopes at $x = 0$.

PROBLEM: Given

- the acceleration as a function of the location $z = f(y)$,

represent

- the velocity as a function of time $v = v(t)$ and
- the location as a function of time $y = y(t)$.

Then our two functions have to satisfy:

$$a_n = f(p_n), \quad v_n = v(t_n), \quad \text{and} \quad p_n = y(t_n).$$

We assume that there is a version of our pair recursive relations,

$$\begin{aligned} v_{n+1} &= v_n + a_n \cdot h \\ p_{n+1} &= p_n + v_n \cdot h \end{aligned}$$

for every $h > 0$ small enough. We substitute these two, as well as $t = t_n$, into our recursive formulas:

$$\begin{aligned} v(t+h) &= v(t) + f(y(t+h)) \cdot h \\ y(t+h) &= y(t) + v(t+h) \cdot h \end{aligned}$$

Then,

$$\begin{aligned} \frac{v(t+h) - v(t)}{h} &= f(y(t+h)) \\ \frac{y(t+h) - y(t)}{h} &= v(t+h) \end{aligned}$$

Taking the limit over $h \rightarrow 0$ gives us the following relations between our functions:

$$\begin{aligned} v'(t) &= f(y(t)), \\ y'(t) &= v(t), \end{aligned}$$

provided $y = y(t)$ and $v = v(t)$ are differentiable at t and $z = f(y)$ is continuous at $y(t)$.
The outcome is treated either as a *system of differential equations of first order* discussed later:

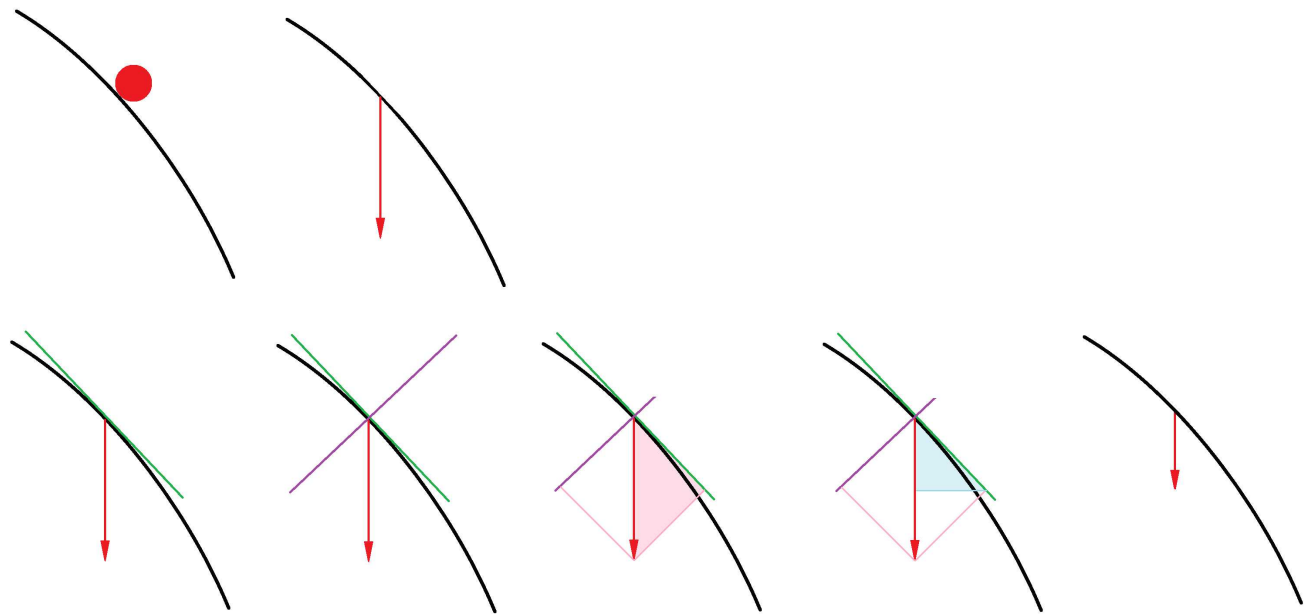
$$\begin{cases} y' &= v, \\ v' &= f(y), \end{cases}$$

which is a vector field, or as a single *differential equation of second order*

$$y'' = f(y) \, .$$

Example 1.14.3: rolling down a slope

The acceleration is not known as a function of time! We just know how it depends on location. As the boll is rolling down a slope, the portion of the gravity acceleration varies:



The increment of the velocity Δv is proportional to the location y and to Δt as in this equation:

$$\Delta v = -f(y) \cdot h, \, f(y) > 0 \, .$$

Then the equation gives us this recursive formula:

$$v(t_{n+1}) = v(t_n) - f(y(t_n)) \cdot h \, .$$

For the location, we have this equation for any interval of time:

$$\Delta y = v \cdot h \, ,$$

where v is known from above. We have a recursive formula:

$$y(t_{n+1}) = y(t_n) + v(t_n) \cdot h \, .$$

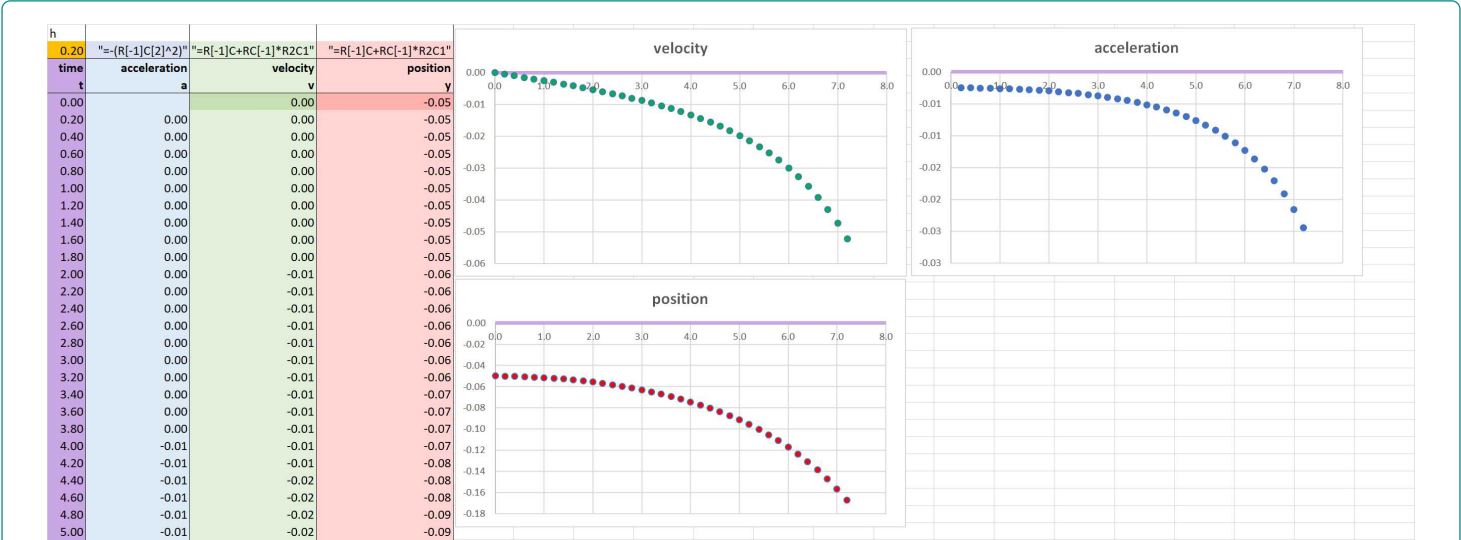
Let's consider

$$f(y) = y^2 \, .$$

For the spreadsheet, we use the following formula that makes a reference to the location:

=-(R[-1]C[2])^2

The result is *accelerated roll*:



Finally, if we think of y and v as sampled differentiable functions, the two equations are converted to ODEs as follows:

$$y' = v, \text{ and } v' = -f(y).$$

Therefore, we have a single ODE:

$$y'' = -f(y).$$

When f is constant, we are back to free fall! A more interesting example is f linear. For example:

$$y'' = y.$$

What are the solutions? First, if $y' = y$, this is also a solution of the above. One class is, therefore:

$$y = Ae^t.$$

More? If $y' = -y$, this is also a solution of the above:

$$y'' = (y')' = (-y)' = -y' = -(-y) = y.$$

Another class is, therefore:

$$y = Be^{-t}.$$

More? After some guessing, we arrive at the *sum* of the above:

$$y = Ae^t + Be^{-t}.$$

This is a *two*-parameter family!

Exercise 1.14.4

What shape of a slide will provide the fastest trip from point A to point B ?

The following is a more general result.

Theorem 1.14.5: General Solution of Roll ODE

The solutions of the ODE:

$$y'' = ky, \; k > 0,$$

are given by:

$$y(t) = Ae^{\sqrt{k}t} + Be^{-\sqrt{k}t},$$

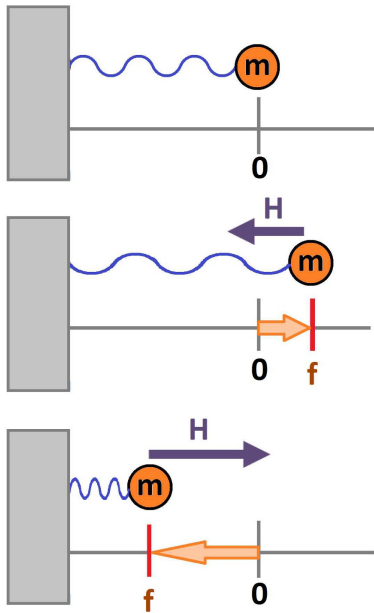
where A, B are two parameters to be determined from the initial conditions.

Exercise 1.14.6

Prove the theorem.

Example 1.14.7: spring

The force of a spring is proportional to the negative of the displacement from the equilibrium.



Description (Hooke’s Law):

- “The acceleration of an object on a spring is proportional to the negative of the current location”.

The increment of the velocity Δv is proportional to the location y and to Δt as in this equation:

$$\Delta v = -k \cdot y \cdot h, \quad k > 0.$$

Suppose we have an initial condition for the velocity:

$$v(t_0) = v_0.$$

Then the equation gives us this recursive formula:

$$v(t_{n+1}) = v(t_n) - ky(t_n) \cdot h.$$

For the location, we have this equation for any interval of time:

$$\Delta y = v \cdot h,$$

where v is known from above. Suppose we also have an initial condition for the location:

$$y(t_0) = y_0.$$

This sets up a recursive formula:

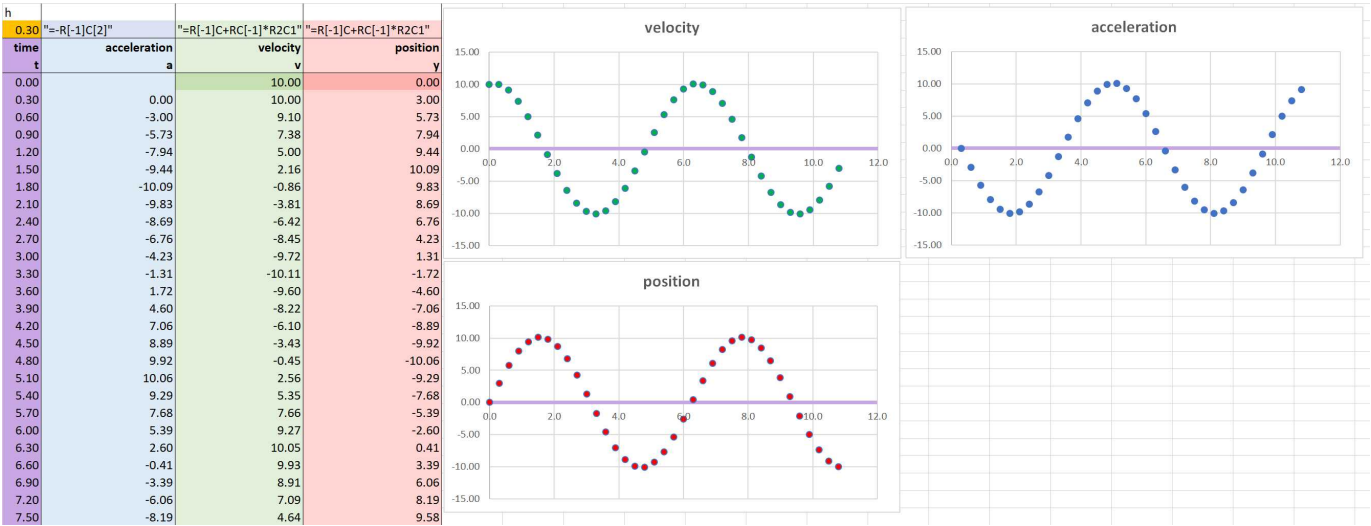
$$y(t_{n+1}) = y(t_n) + v(t_n) \cdot h.$$

The combination of the two recursive formulas gives us a discrete model.

We modify the spreadsheet by giving the acceleration a formula that makes a reference to the location:

```
=-R[-1]C[2]
```

The result is *oscillation*:



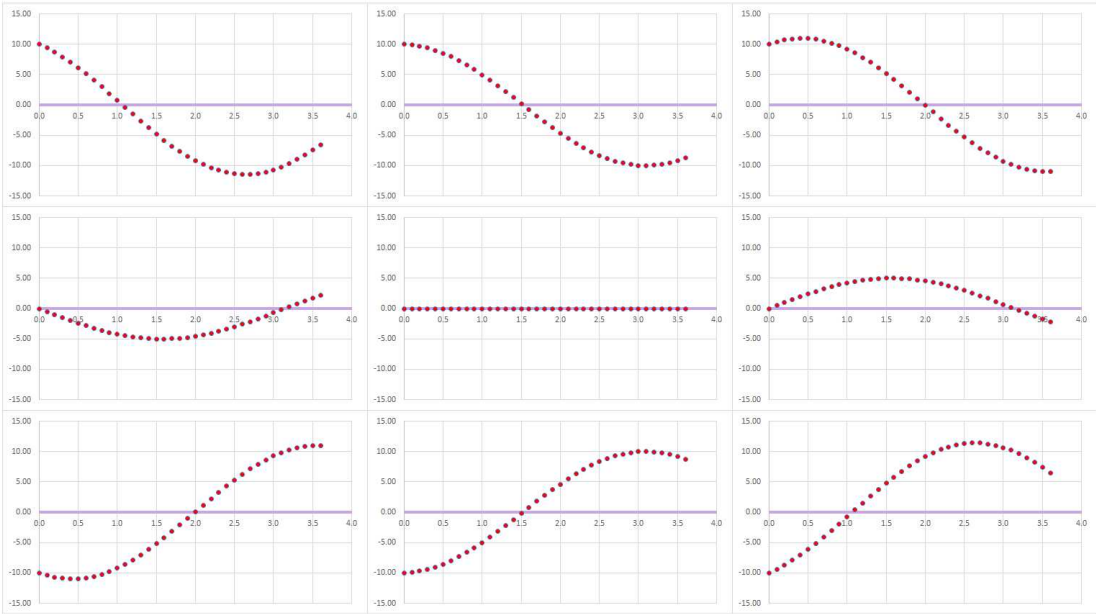
Finally, if we think of y and v as sampled differentiable functions, the two equations are converted to ODEs as follows:

$$y' = v, \text{ and } v' = -ky.$$

Therefore, we have a single ODE:

$$y'' = -ky, \text{ } k > 0.$$

This is a sketch of the solution set:



When $k = 1$, a couple of solutions of $y'' = -y$ are worth remembering:

$$y_1(t) = \cos t \text{ and } y_2(t) = \sin t.$$

More solutions? After some guessing, we arrive at the *linear combination* of the above:

$$y = A \cos t + B \sin t.$$

This is a *two*-parameter family!

The following is a more general result.

Theorem 1.14.8: General Solution of Spring ODE

The solutions of the ODE:

$$y'' = -ky, \text{ } k > 0,$$

are given by:

$$y(t) = A \cos \sqrt{k}t + B \sin \sqrt{k}t$$

where A, B are two parameters to be determined from the initial conditions.

Exercise 1.14.9

Prove the theorem.

Another example of such motion is the pendulum (Chapter 4). For more examples of wave functions, see Chapter 1PC-5.

Example 1.14.10: spring with dampening

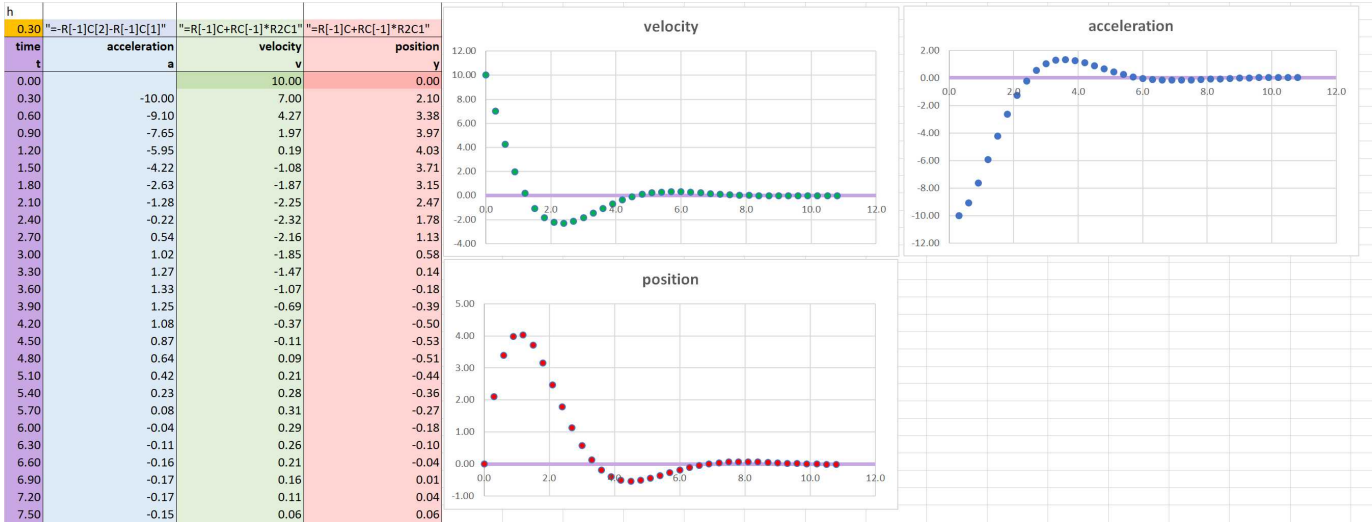
The acceleration can also depend on the velocity! For example, one might want to stop the up and down motion of the suspension of a car. How? By introducing some *friction*. A model of friction may be that its force is proportional to the speed of the motion but in the opposite direction, i.e., proportional to the negative of the velocity:

$$-my'.$$

We just modify the spreadsheet by giving the acceleration a formula that makes a reference to both the location and the velocity:

$$=-R[-1]C[2]-R[-1]C[1]$$

The result is a quick disappearance of the oscillation:



The ODE is as follows:

$$y'' = -my' - ky, \quad m, k > 0.$$

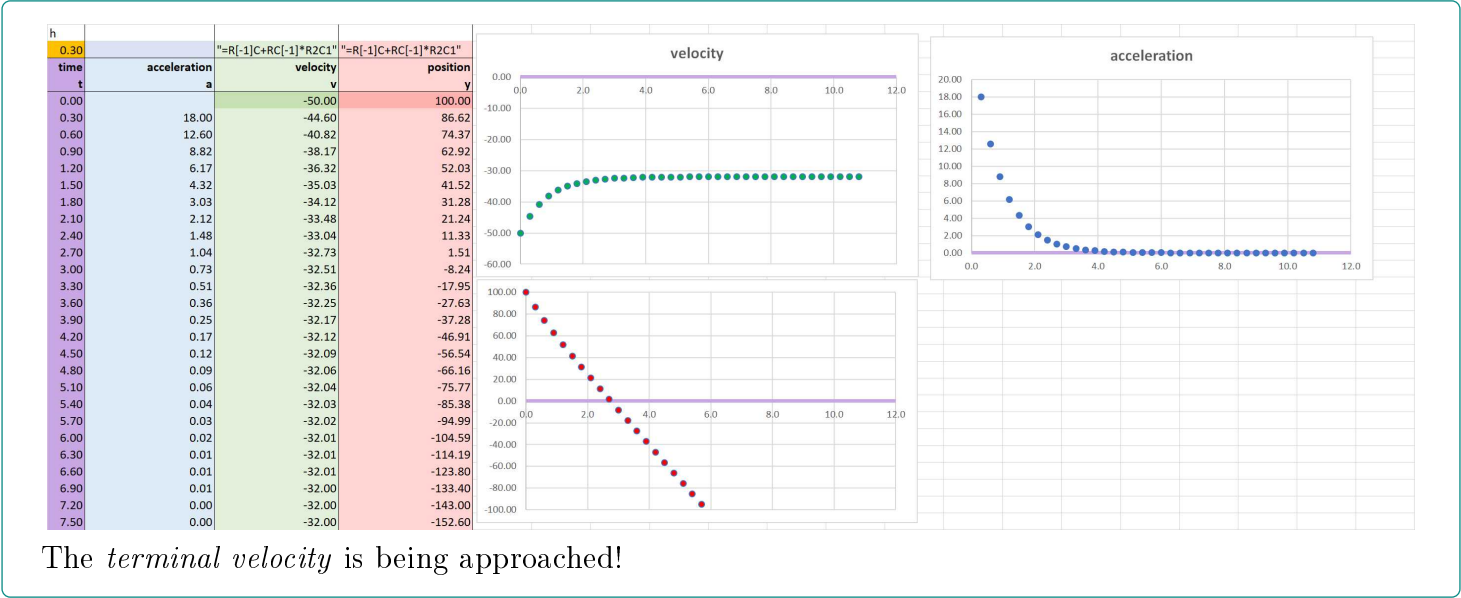
It needs further analysis.

Example 1.14.11: free fall with air resistance

The effect of air resistance on a falling ball is the same:

$$y'' = -my' - g.$$

These are the results of simulations:



Definition 1.14.12: ODEs of second order

Suppose h is some function of three variables. Then a *difference equation of second order* is defined to be the following:

$$\frac{\Delta^2 y}{\Delta t^2} = h\left(t, y, \frac{\Delta y}{\Delta t}\right)$$

and an *ordinary differential equation of second order* is defined to be the following:

$$\frac{d^2 y}{dt^2} = h\left(t, y, \frac{dy}{dt}\right) \quad \text{or} \quad y'' = h(t, y, y')$$

It is more productive to address these equations as *systems of ODEs* of first order (Chapter 3):

$$\begin{cases} y' = v \\ v' = h(t, y, v) \end{cases}$$

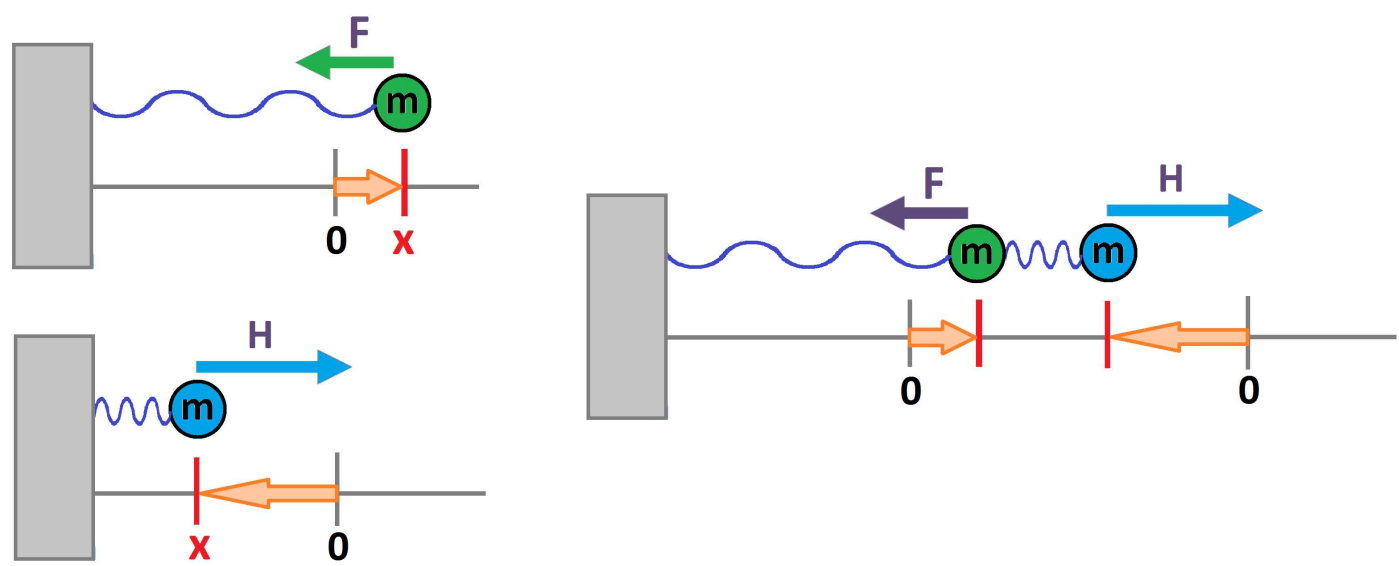
Definition 1.14.13: linear ODEs of second order

These ODEs are called *linear* when h is linear:

$$ay'' + by' + cy = 0, \quad a \neq 0.$$

Almost all examples have been linear. They have special properties.

For example, attaching to our spring another, identical spring will create an oscillation equivalent to that of a single spring:



In other words, the sum of two solutions is also a solution!

More general is the following.

Theorem 1.14.14: Linearity of 2nd Order Linear ODE

If y_1 and y_2 are solutions of a 2nd Order Linear ODE, then so is any of their linear combinations; i.e., if y_1 and y_2 satisfy

$$y'' + by' + cy = 0,$$

then so does

$$y = Ay_1 + By_2,$$

where A, B are any real numbers.

Proof.

It follows from linearity of differentiation.

The result is also known as the “Superposition Principle”.

In two theorems, we showed that formulas for the solutions of the Roll ODE,

$$y'' = ky \ (k > 0),$$

the solutions of the Spring ODE,

$$y'' = -ky \ (k > 0),$$

satisfy the equation by direct substitutions. But how would one *discover* them?

The idea of the former is that the solution is an exponential function:

$y = e^{rt}.$

We substitute:

$$(e^{rt})'' = ke^{rt}.$$

Compute:

$$r^2e^{rt} = ke^{rt}.$$

We discover that

$$r = \pm\sqrt{k}.$$

The result matches the theorem. If we apply this idea to the latter, we arrive to:

$$r = \pm \sqrt{-k}.$$

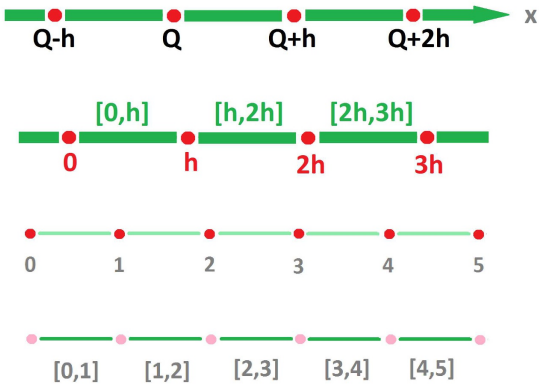
These are imaginary numbers! How imaginary (and complex) numbers produce sine and cosine is shown in the next chapter.

1.15. Discrete forms, continued

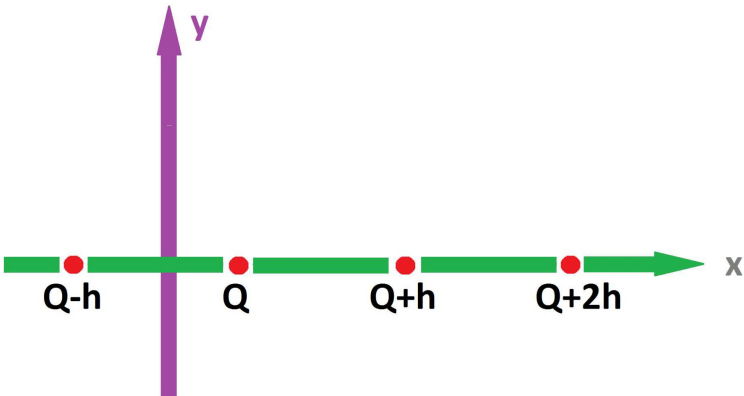
Let’s review.

We divide the x -axis (i.e., the real line \mathbf{R}) into discrete pieces into intervals of equal length $h > 0$ starting from some location Q . These are the two types of pieces:

- The *nodes* are 0-cells, $x = \dots, Q - 2h, Q - h, Q, Q + h, Q + 2h, \dots$
- The *edges* are the 1-cells, $[x, x + h] = \dots [Q - h, Q], [Q, Q + h], [Q + h, Q + 2h], \dots$



In the meantime, the y -axis is just the reals:

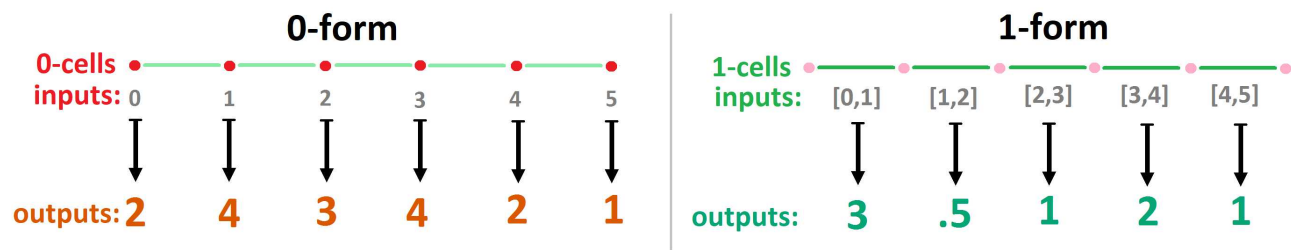


Now, the main objects in our study will be these two types of discrete functions:

- The functions that have nodes as inputs are called 0-forms.
- The functions that have edges as inputs are called 1-forms.

The outputs are real numbers.

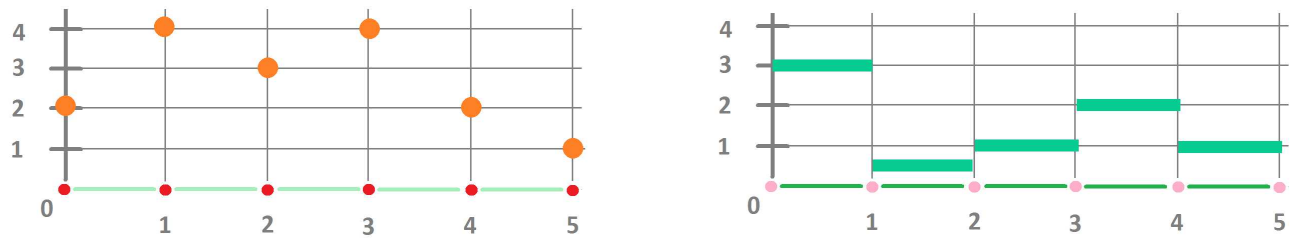
We use arrows to picture these functions as correspondences:



We can *list the values* of the two functions:

- a 0-form (node function) f with $f(0) = 2$, $f(1) = 4$, $f(2) = 3$, ...
- a 1-form (edge function) s with $s([0, 1]) = 3$, $s([1, 2]) = .5$, $s([2, 3]) = 1$, ...

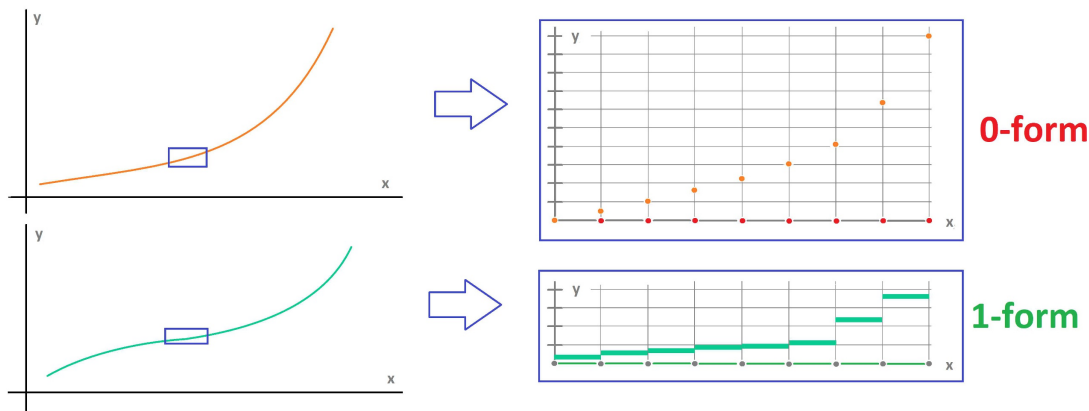
Given a function f , its graph is the collection of points on the xy -plane so that we always have $y = f(x)$:



Above we see the following:

1. For a node function, x is a node, a number, and $y = f(x)$ is also a number. Together, they produce (x, y) , a point on the xy -plane (with the x -axis split into cells as shown above).
2. For an edge function, $[A, B]$ is an edge, an interval in the x -axis, and $y = g([A, B])$ is a number. Together, they produce a collection of points on the xy -plane such as (x, y) for every x in $[A, B]$. The result is a horizontal segment.

We might discover such functions if we zoom in on a *continuous curve*:



Now a new question: What can we say about *the change of the change*?

For example, we progress from the locations defined at the nodes of the partition of the time line to the changes of locations (displacements) defined on the edges to the changes of displacements defined at the nodes again:

f :	—	2	— — —	5	— — —	10	—
Δf :	—	— ● —	5 — 2 = 3	— ● —	10 — 5 = 5	— ● —	—
$\Delta \Delta f$:	—	— ? —	— — —	5 — 3 = 2	— — —	— ? —	—
t :		1	2	3	4	5	

Generally, if we know only *three* values of a function (first line) at the ends of an interval, we compute the

differences along the two intervals (second line) and place the results at the corresponding edge:

—	$f(x_1)$	— — —		$f(x_2)$	— — —	$f(x_3)$	—
—	— • —	$\Delta\left(f[x_1, x_2]\right)$		— • —	$\Delta f\left([x_2, x_3]\right)$	— • —	—
—	— • —	— — —	$\Delta f\left([x_2, x_3]\right) - \Delta\left(f[x_1, x_2]\right)$	— — —	— • —	—	—
	x_1			x_2		x_3	

To find the change of this new function, we carry out the same operation and place the result in the middle (third line).

What can we say about the *rate of change of the rate of change*? Add division.

For example, we progress from the locations defined at the nodes of the partition of the time line to the velocities defined on the edges to the accelerations defined at the nodes again:

location:	—	2	— — —	5	— — —	10	—
velocity:	—	— • —	$\frac{5-2}{2} = 3/2$	— • —	$\frac{10-5}{2} = 5/2$	— • —	—
acceleration:	—	— ? —	— — —	$\frac{5/2 - 3/2}{2} = 1/2$	— — —	— ? —	—
time:		1	2	3	4	5	

Generally, if we know only *three* values of a function (first line) at the ends of an interval, we compute the difference quotients along the two intervals (second line) and place the results at the corresponding edge:

—	$f(x_1)$	— — —	$f(x_2)$	— — —	$f(x_3)$	—
—	— • —	$\frac{\Delta f}{h}$	— • —	$\frac{\Delta f}{h}$	— • —	—
—	— • —	— — —	$\frac{\frac{\Delta f}{h} - \frac{\Delta f}{h}}{h}$	— — —	— • —	—
	x_1		x_2		x_3	

To find the rate of change of this new function, we carry out the same operation and place the result in the middle (third line)

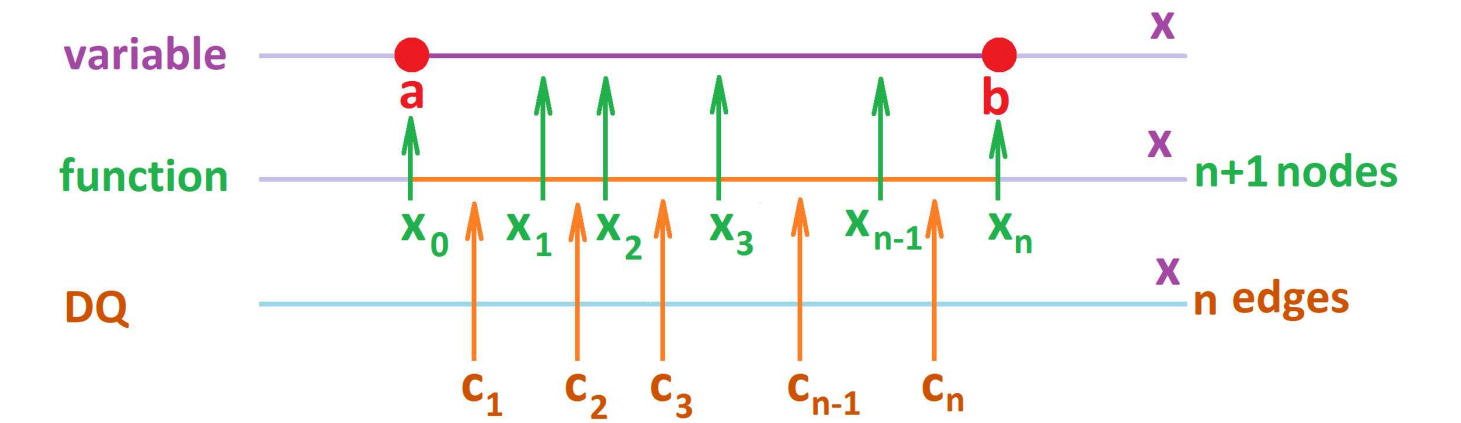
Let’s review the construction of the difference quotient in full generality.

First, we have a *cell decomposition* of an interval $[a, b]$. We partition it into n edges with the help of the nodes:

$$a = x_0, \, x_1, \, x_2, \, ..., \, x_{n-1}, \, x_n = b.$$

These are the edges:

$$c_1 = [x_0, x_1], \, c_2 = [x_1, x_2], \, ..., \, c_n = [x_{n-1}, x_n].$$



If a function $y = f(x)$ is defined at the nodes $x_k, \, k = 0, 1, 2, ..., n$, the *difference* of f is defined at the edges by:

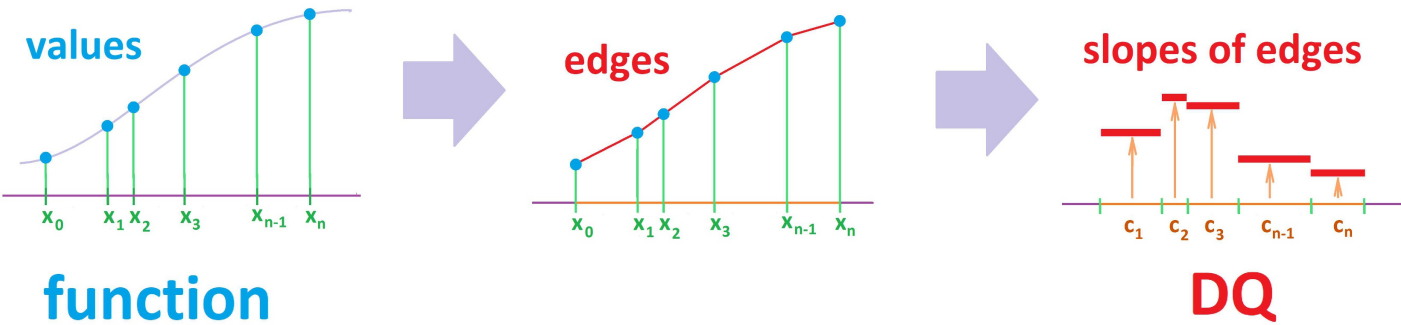
$$\Delta f(c_k) = f(x_{k+1}) - f(x_k)$$

for each $k = 1, 2, \dots, n$. Also, the *difference quotient* of f is defined at the edges:

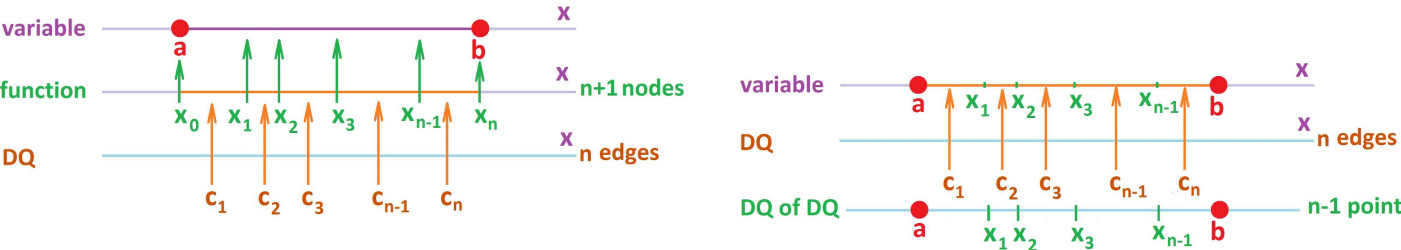
$$\frac{\Delta f}{\Delta x}(c_k) = \frac{f(x_{k+1}) - f(x_k)}{x_{k+1} - x_k}$$

for each $k = 1, 2, \dots, n$.

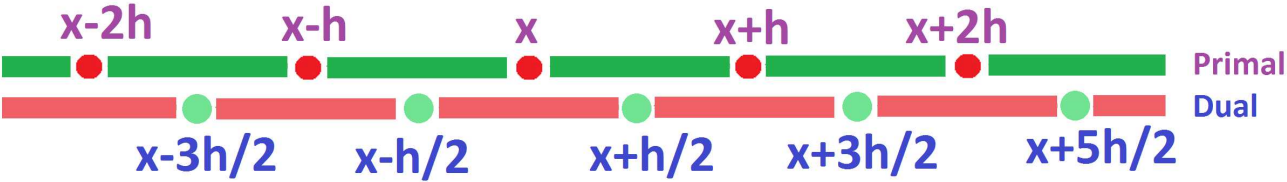
The former function represents the change of value from node to node, while the latter function represents the slopes of the secant lines over the nodes of the decomposition:



In order to repeat the difference or difference quotient construction on this new function, we will need now a new *decomposition*:



We need convert edge functions to node functions and vice versa by switching between nodes and edges. To set this up we use the following match between the cells:



Thus there are a new node for each edge in the domain and a new edge for each node. Together, these new nodes and edges form a new copy of the domain.

This is how the cells above are matched above:

- An edge $[x, x + h]$ corresponds to the node $x + h/2$.
- A node x corresponds to the edge $[x - h/2, x + h/2]$.

For the general situation, we have the following.

Definition 1.15.1: primal and dual domains

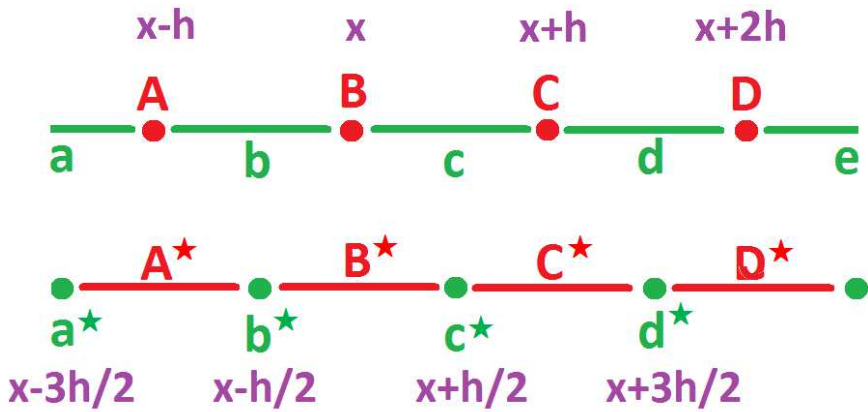
A cell decomposition of a copy of the real line \mathbf{R} is called the *primal domain*. A cell decomposition of another copy of the real line \mathbf{R} is called the *dual domain*. A function from 0-cells of the former to the 1-cells of the latter and from 1-cells of the former to the 0-cells of the latter that is monotone is called the *star operator* or the \star -operator.

This is how they are denoted:

Duality of cells

A primal cell a corresponds to a dual cell a^* .

This is how they match up:



The relation can be reversed:

- A dual edge $a = [x - h/2, x + h/2]$ corresponds to the primal node $a^* = x$.
- A dual node $x + h/2$ corresponds to the primal edge $x^* = [x, x + h]$.

Definition 1.15.2: dual cells

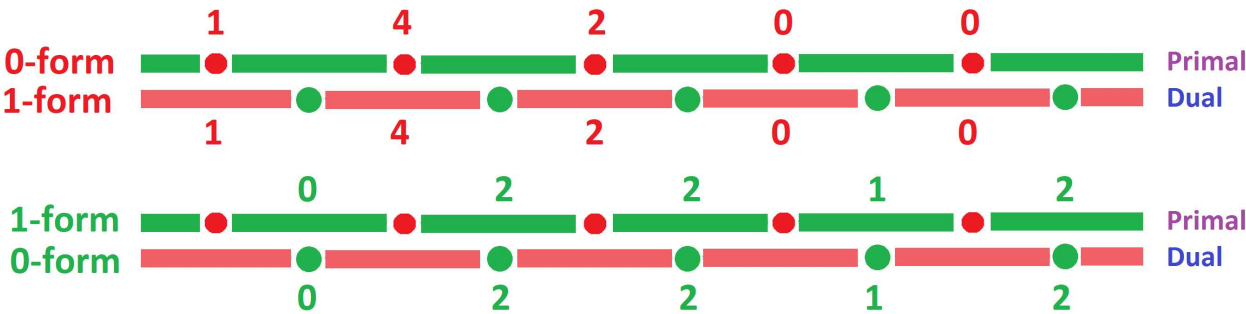
For both primal and dual and for both nodes and edges, a and a^* are called *dual* of each other.

In other words, we have:

- A primal k -cell corresponds to a dual $(1 - k)$ -cell.

Now the *duality of forms*.

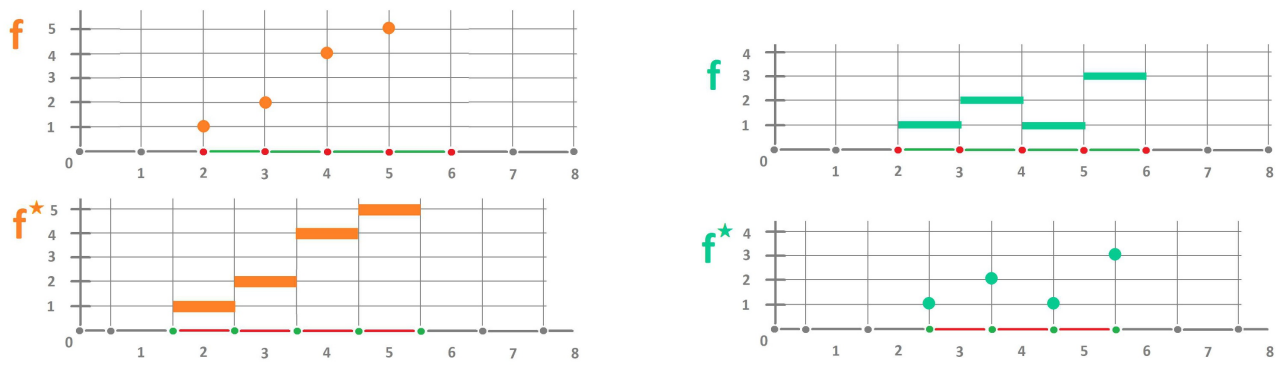
The values are the same, only the inputs change! Two dual pairs of forms are illustrated below:



They come in pairs, primal and dual:

- If f is a primal node function, a 0-form, then $f^*\left([x - h/2, x + h/2]\right) = f(x)$ is a dual edge function, a 1-form.
- If g is a primal edge function, a 1-form, then $g^*(x) = g\left([x - h/2, x + h/2]\right)$ is a dual node function, 0-form.

These are the graphs of the dual forms:



If we zoom out, the functions dual to each other will look identical!

The following is the general approach.

Definition 1.15.3: dual forms

Given a node/edge function s , the following is an edge/node function:

$$s^*(a) = s(a^*)$$

Then, s^* is called the *dual function* of s .

This setup allows us to construct the second difference of a 0-form.

We partition the segment into $n - 1$ intervals by giving nodes to the edges of the last decomposition with the same names:

$$p = c_1, \ c_2, \ c_3, \ ..., \ c_{n-1}, \ c_n = q \ .$$

Then the increments are:

$$\Delta c_k = c_{k+1} - c_k \ .$$

Now, what are the nodes corresponding to the edges of this new decomposition? The nodes of the last decomposition of course! Indeed, we have:

$$x_1 \text{ in } [c_1, c_2], \ x_2 \text{ in } [c_2, c_3], \ ..., \ x_{n-1} \text{ in } [c_{n-1}, c_n] \ .$$

We apply the same constructions to this decomposition to the function $g = \frac{\Delta f}{\Delta x}$. The difference function of g is defined at the edges of the new decomposition by:

$$\Delta g(x_k) = g(c_{k+1}) - g(c_k)$$

for each $k = 1, 2, ..., n - 1$. The difference quotient function of g is defined at the edges of the new decomposition by:

$$\frac{\Delta g}{\Delta x}(x_k) = \frac{g(c_{k+1}) - g(c_k)}{c_{k+1} - c_k}$$

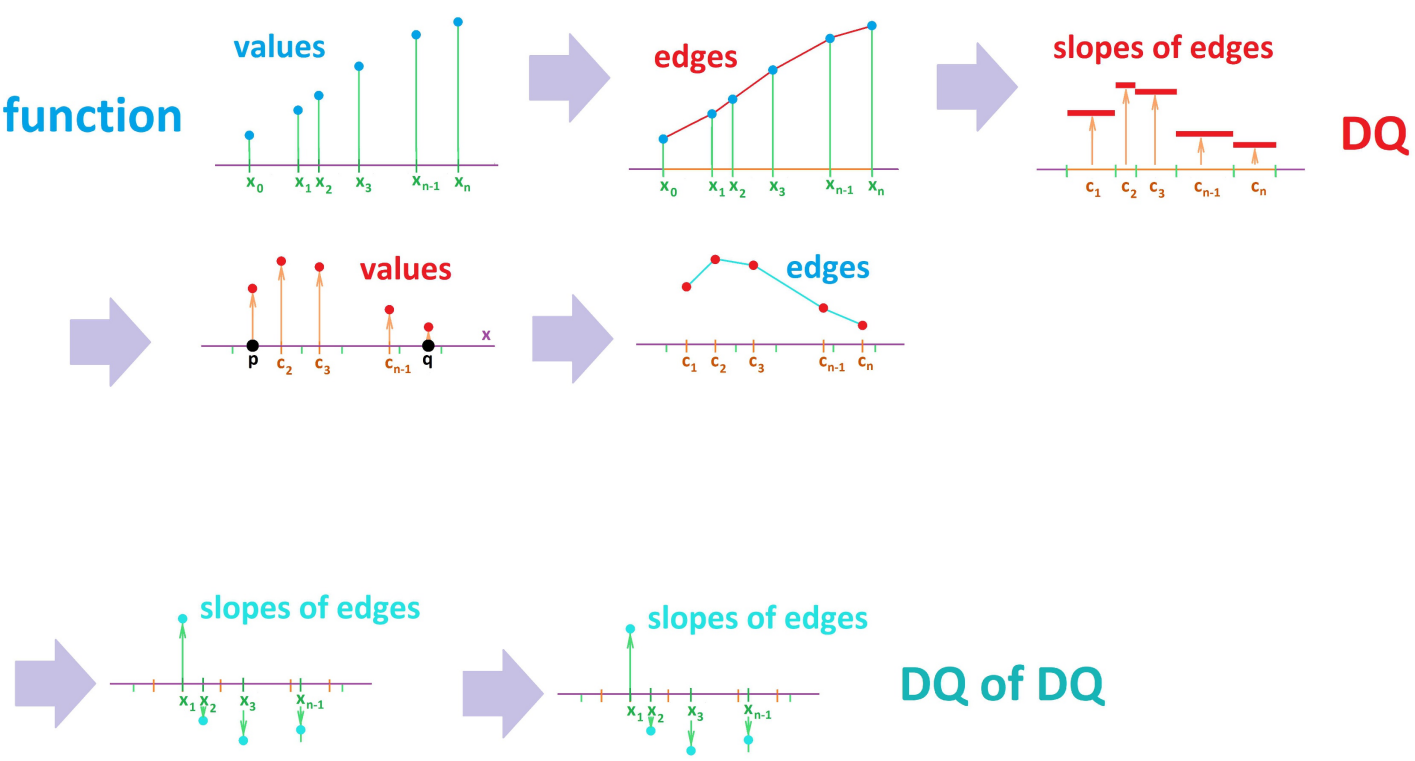
for each $k = 1, 2, ..., n - 1$.

Definition 1.15.4: second difference

The *second difference* of a function f is defined to be the difference of the difference, i.e., it is defined at the nodes of the decomposition (and denoted) as follows:

$$\Delta^2 f(x_k) = \Delta f(c_{k+1}) - \Delta f(c_k)$$

for each $k = 1, 2, \dots, n - 1$.



Definition 1.15.5: second difference quotient

The *second difference quotient* of a function f is defined to be the difference quotient of the difference quotient, i.e., it is defined at the nodes of the decomposition (and denoted) as follows:

$$\frac{\Delta^2 f}{\Delta x^2}(x_k) = \frac{\frac{\Delta f}{\Delta x}(c_{k+1}) - \frac{\Delta f}{\Delta x}(c_k)}{c_{k+1} - c_k}$$

for each $k = 1, 2, \dots, n - 1$.

Note that there are:

- $n + 1$ values of f (at the nodes),
- n values of $\frac{\Delta f}{\Delta x}$ (at the edges), and
- $n - 1$ values of $\frac{\Delta^2 f}{\Delta x^2}$ (at the nodes except a and b).

We will often omit the subscripts for the simplified notation:

Second difference

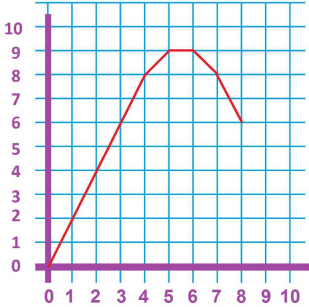
$$\Delta^2 f(x) = \Delta f(c + \Delta c) - \Delta f(c) .$$

Second difference quotient

$$\frac{\Delta^2 f}{\Delta x^2}(x) = \frac{\frac{\Delta f}{\Delta x}(c + \Delta c) - \frac{\Delta f}{\Delta x}(c)}{\Delta c}.$$

Example 1.15.6: curvature

As we know, the difference quotient of a linear function is constant. The second difference quotient is, therefore, zero. We conclude that a non-zero second difference quotient indicates a non-linear graph:



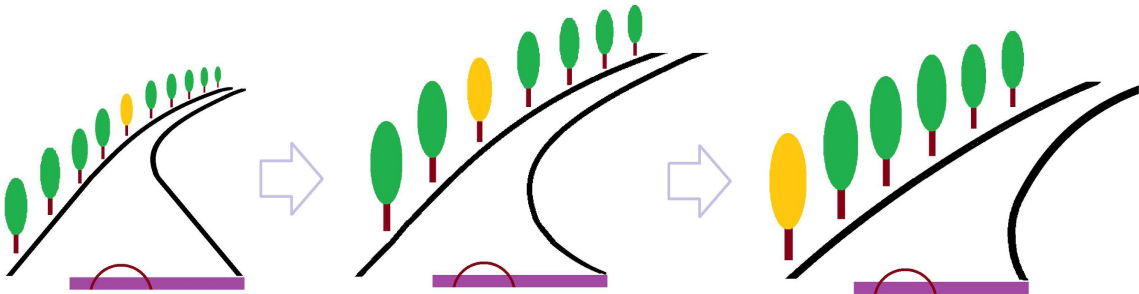
Above, the slopes remain the same, 2, at first; therefore, the second difference quotient is zero:

$$\frac{\Delta^2 f}{\Delta x^2} = 0.$$

Then, the slope changes to 1 and this change is the second difference quotient (assuming $\Delta x = 1$):

$$\frac{\Delta^2 f}{\Delta x^2} = -1.$$

As another way to see this idea, imagine yourself driving along a straight part of the road and seeing a particular tree ahead (no curvature), then, as you start to turn, the trees start to pass your field of vision from right to left (curvature):



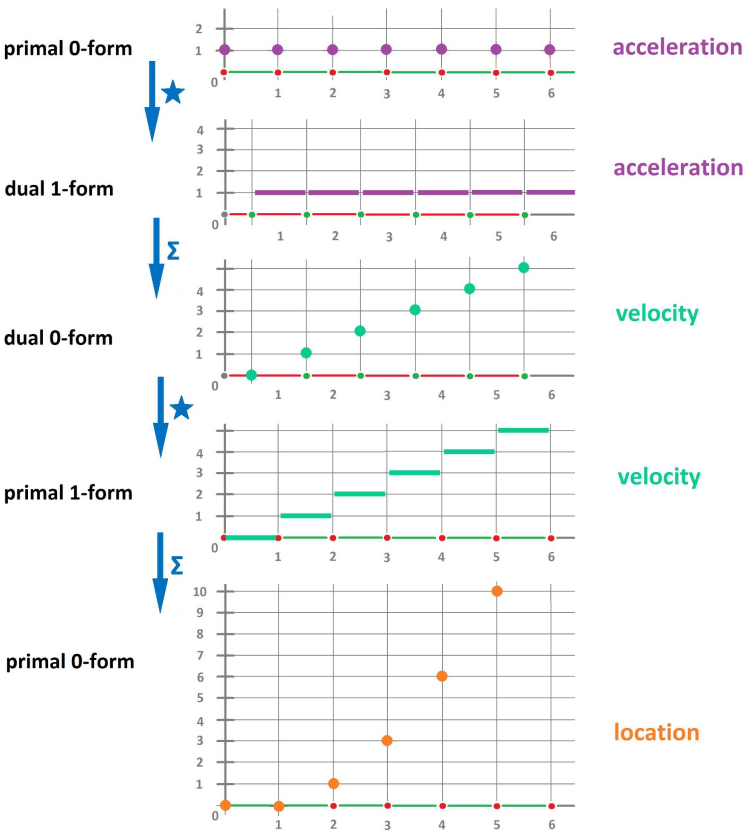
Furthermore, higher values of the *curvature* of the graph of $y = f(x)$. The higher value of the second difference quotient means higher values of the *curvature* of the graph of $y = f(x)$.

Example 1.15.7: free fall

Description: “The acceleration is constant”. We have a difference equation of second order:

$$\Delta^2 p = a(\Delta t)^2.$$

This is just a new representation of the same discrete model we have used before. To illustrate, let’s try this specific choice of $a = 1$ and $\Delta t = 1$:



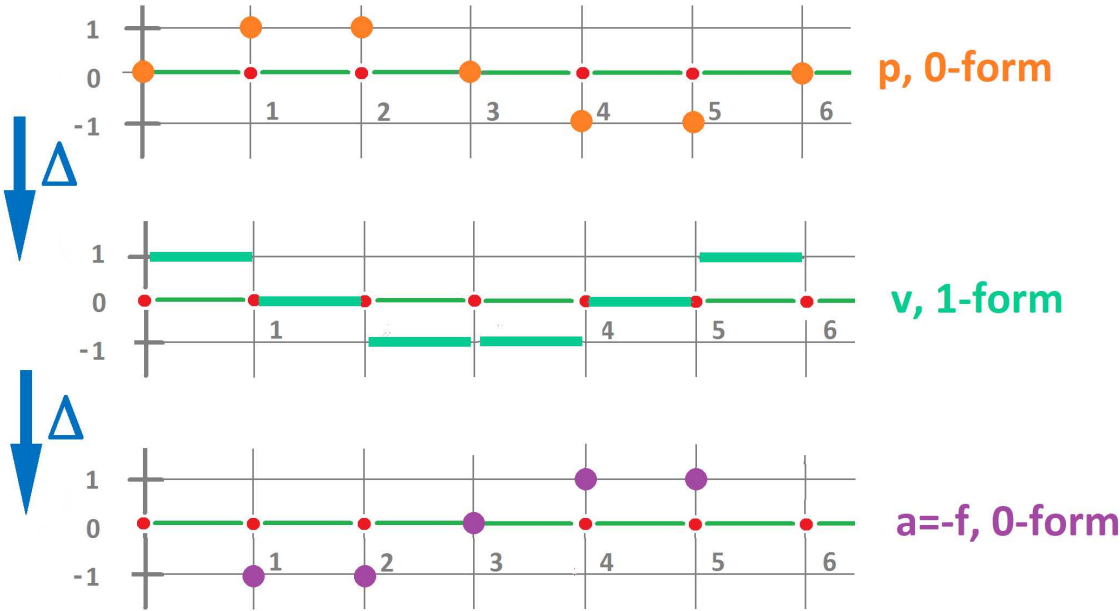
We start with a constant function, do summation twice, and arrive to a function that seems quadratic.

Example 1.15.8: oscillating spring

Description: “The acceleration of an object on a spring is proportional to the negative of the current location”. We have a difference equation of second order for the position p of the end of the spring:

$$\Delta^2 p = -kp(\Delta t)^2.$$

This is just a new representation of the same discrete model we have used before. To illustrate, let’s try this specific choice of a periodic p with $k = 1$ and $\Delta t = 1$:



After taking two differences (the star operators are omitted), we arrive to $-p$. A match!

Exercise 1.15.9

Illustrate similarly other models from this chapter.

Chapter 2: Vector variables

Contents

2.1 Where matrices come from	126
2.2 Transformations of the plane	133
2.3 Linear operators	141
2.4 Examining and building linear operators	149
2.5 The determinant of a matrix	162
2.6 It's a stretch: eigenvalues and eigenvectors	168
2.7 The significance of eigenvectors	179
2.8 Bases	184
2.9 Classification of linear operators according to their eigenvalues	191

2.1. Where matrices come from

In this chapter, we set aside the geometry of the Euclidean spaces – measuring distances and angles – and concentrate on *pure algebra*.

Let's recall these problems about *mixtures*.

Problem, dimension 1: Suppose we have a type of coffee that costs \$3 per pound. How much do we get for \$60?

Let x be the weight of the coffee. Then we have:

Setup: $3x = 60$. Solution: $x = \frac{60}{3}$.

Problem, dimension 2: Suppose the Kenyan coffee costs \$2 per pound and the Colombian coffee costs \$3 per pound. How much of each do you need to have 6 pounds of blend with a total price of \$14?

Let x be the weight of the Kenyan coffee and let y be the weight of Colombian coffee. Then we have:

Setup:
$$\begin{array}{rcl} x & + & y & = & 6 \\ 2x & + & 3y & = & 14 \end{array}$$

Solution: From the first equation, we derive: $y = 6 - x$. Then substitute it into the second equation: $2x + 3(6 - x) = 14$. Solve the new equation: $-x = -4$, or $x = 4$. Substitute this back into the first equation: $(4) + y = 6$, then $y = 2$.

But it was so much simpler for the former problem! Is it possible to mimic the setup, i.e., the equation, and the solution of the 1-dimensional case for the 2-dimensional case? The existence of vector algebra suggests that it might be possible.

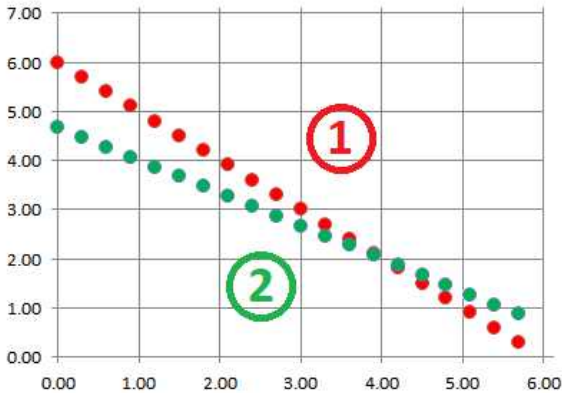
Let’s recall the ways we have interpreted the problem.

First: points and lines.

We think of the two equations as equations about the coordinates of points, (x,y) , in the plane:

$$\begin{cases} x + y = 6, \\ 2x + 3y = 14. \end{cases}$$

Either equation is a line on the plane. The solution $(x,y) = (4,2)$ is the point of their intersection:



Second: vectors and their linear combinations.

Let’s put the equations in these tables:

$1 \cdot x$	+	$1 \cdot y$	=	6
$2 \cdot x$	+	$3 \cdot y$	=	14

The table is split horizontally to reveal the equations. Next, we start to split vertically and realize that we see a componentwise *addition of vectors*:

1	·	x	+	1	·	y	=	6
2	·	x	+	3	·	y	=	14

We have:

1	·	x		1	·	y	=	6
2	·	x	+	3	·	y	=	14

But x ’s and y ’s are repeated! We realize that we see a componentwise *scalar multiplication of vectors*:

1		·	x	+	1		·	y	=	6
2		·	x	+	3		·	y	=	14

Vectors start to appear. Indeed, our system has been reduced to a single *vector* equation:

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} x + \begin{bmatrix} 1 \\ 3 \end{bmatrix} y = \begin{bmatrix} 6 \\ 14 \end{bmatrix}$$

We see three vectors of the same dimension 2. This isn’t a coincidence. They are of the same nature:

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 6 \\ 14 \end{bmatrix} \text{ are } \begin{bmatrix} \text{weight (in pounds)} \\ \text{cost (in dollars)} \end{bmatrix}.$$

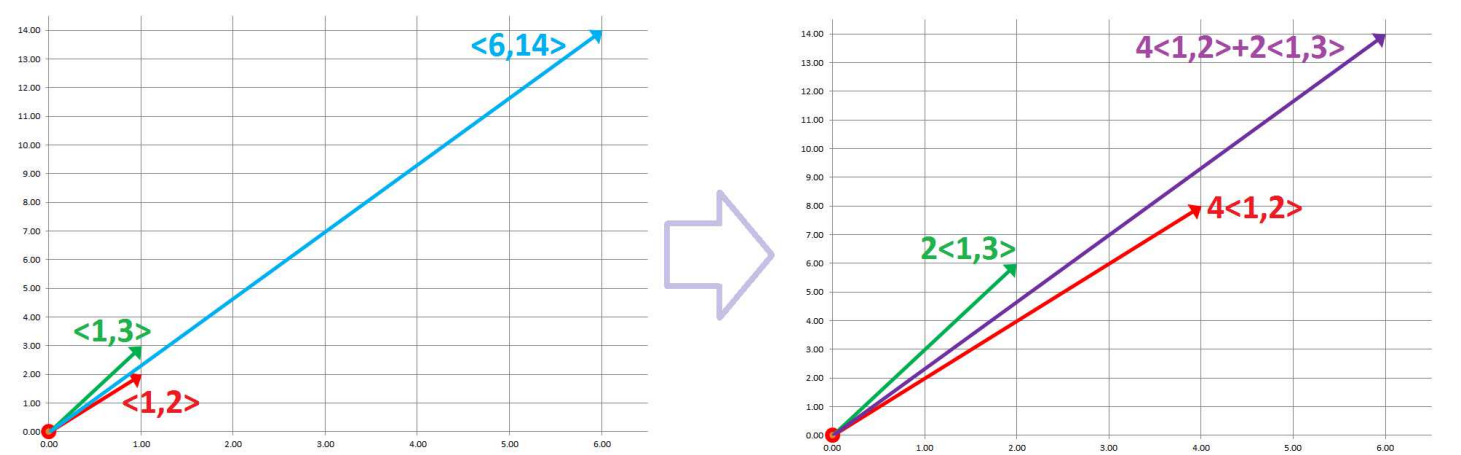
They live in the same space \mathbf{R}^2 and, therefore, subject to the operations of vector algebra.

The solution to the system in this interpretation has the following algebraic meaning. We can think of the two equations as a single equation about the coefficients, x and y , of these vectors in the plane:

$$x \begin{bmatrix} 1 \\ 2 \end{bmatrix} + y \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 6 \\ 14 \end{bmatrix} .$$

Geometrically, we need to find a way to stretch these two vectors so that after adding them the result is the vector on the right. We speak of *linear combinations*.

The setup is on the left followed by a trial-and-error on the right:



So, the new point of view has changed: Instead of the *locations*, we are after the *directions*.

Exercise 2.1.1

Are there other vectors here?

Third: transformations.

Initially, we use points.

Dimension 1 problem:

- A transformation $f : \mathbf{R} \rightarrow \mathbf{R}$ is given by

$$f(x) = 30x .$$

- Solve the equation:

$$f(x) = 60 .$$

Dimension 2 problem:

- A transformation $f : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ is given by

$$F(x, y) = (x + y, 2x + 3y) .$$

- Solve the equation:

$$F(x, y) = (6, 14) .$$

Now, we prefer to use vectors. Let’s find all 2-dimensional vectors in the equations:

$$\begin{array}{rcl} 1 \cdot x & + & 1 \cdot y = 6 \\ 2 \cdot x & + & 3 \cdot y = 14 \end{array}$$

The first is on the right; it consists of the two “free” terms (free of x ’s and y ’s!) on the right-hand side:

$$B = \begin{bmatrix} 6 \\ 14 \end{bmatrix} .$$

Another one is less visible; it is made up of the two unknowns:

$$X = \begin{bmatrix} x \\ y \end{bmatrix} .$$

Even though its dimension is also 2, it's not of the same nature as the others:

$$\begin{bmatrix} x \\ y \end{bmatrix} \text{ is } \begin{bmatrix} \# \text{ of pounds} \\ \# \text{ of pounds} \end{bmatrix} .$$

It lives in a different \mathbf{R}^2 .

Then, we have a function between these two spaces:

$$F : \mathbf{R}^2 \rightarrow \mathbf{R}^2, \quad Y = F(X) .$$

Its formula can be written in terms of vectors:

$$F \left(\begin{bmatrix} x \\ y \end{bmatrix} \right) = \begin{bmatrix} x + y \\ 2x + 3y \end{bmatrix} .$$

Our problem becomes a problem of solving an equation for X :

$$F(X) = B .$$

Warning!

Since we aren't doing any geometry but we are doing vector algebra, the vector approach will be preferred throughout the chapter.

The algebraic operations needed to compute F are so simple that they will be easy to abbreviate.

Let's review the setup. The problem for dimension n has n ingredients:

	dim 1	dim 2
the unknown	x	$X = \langle x, y \rangle$
multiplied by	3	?
is equal to	60	$B = \langle 6, 14 \rangle$

We have transitioned from numbers to vectors. But what is the operation that makes B from X ? None of the familiar ones.

The four coefficients of x, y form a table:

$$F = \begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix} .$$

It has two rows and two columns. In other words, this is a 2×2 *matrix*.

Both X and B are *column-vectors* in dimension 2, and matrix F turns X into B . This is very similar to multiplication of numbers; after all, they are vectors of dimension 1. Let's match the setups of the two problems:

dim 1 : $m \cdot x = b$

dim 2 : $F \cdot X = B$

If we can just make sense of the new algebra!

Here $FX = B$ is a *matrix equation*, and it's supposed to capture the system of equations. Let's compare the original system of equations to $FX = B$:

$$\begin{array}{rcl} x & +y & = 6 \\ 2x & +3y & = 14 \end{array} \text{ , rewritten as } \begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 6 \\ 14 \end{bmatrix} .$$

We can see these equations on the right-hand side as these two *dot products*. First:

$$1 \cdot x + 1 \cdot y = 6 \text{ , rewritten as } \begin{bmatrix} 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = 6 \text{ .}$$

Second:

$$2x + 3y = 14 \text{ , rewritten as } \begin{bmatrix} 2 & 3 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = 14 \text{ .}$$

This suggests what the meaning of FX should be. We “multiply” either row in A , as a vector, by the vector X – via the dot product:

Definition 2.1.2: product of matrix and vector

The *product* FX of a 2×2 matrix F and a 2-vector X ,

$$F = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad X = \begin{bmatrix} x \\ y \end{bmatrix}$$

is defined to be the following 2-vector:

$$FX = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}$$

We can still see these dot products in the result:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix} \text{ .}$$

Warning!

A matrix is nothing but an abbreviation of a transformation of the plane:

$$FX = F(X) \text{ .}$$

However, not all transformations can be represented by matrices.

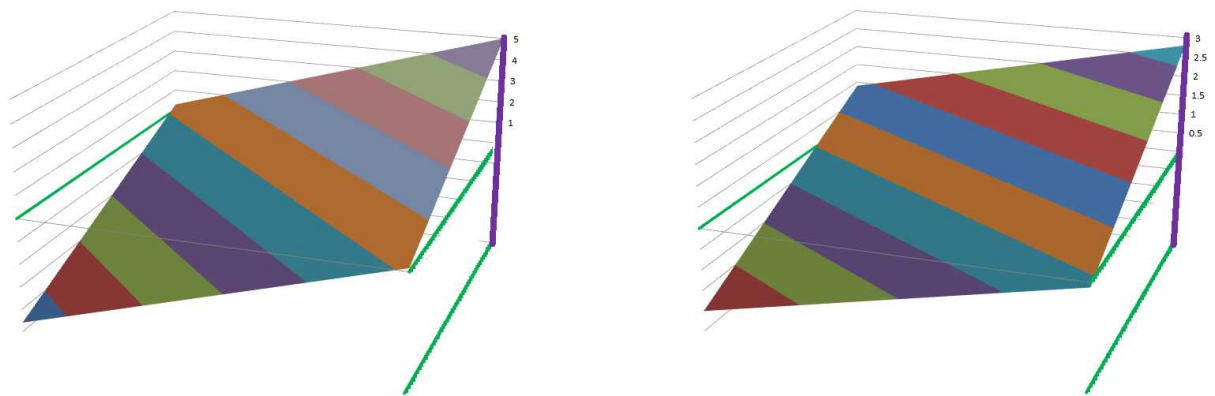
Example 2.1.3: 3 variables

What if the blend is to contain another, third, type of coffee? Given three prices per pound, 2, 3, 5, how much of each do you need to have 6 pounds of blend with a total price of 14?

Let x , y , and z be the weights of the three types of coffee, respectively. Then the total price of the blend is 14. Therefore, we have a system:

$$\begin{cases} x + y + z = 6 \\ 2x + 3y + 5z = 14 \end{cases}$$

Either of these equations represents a plane in \mathbf{R}^3 . The solution set then comes from their intersection:



There are, of course, infinitely many solutions. An additional restriction in the form of another linear equation may reduce the number to one... or none. The variety of possible outcomes is, by far, higher than in the 2-dimensional case; they are not discussed in this chapter.

The vector algebra, however, is the same! The three weights can be written in a vector, $\langle 1, 1, 1 \rangle$, and the first equation becomes the dot product:

$$\langle 1, 1, 1 \rangle \cdot \langle x, y, z \rangle = 6.$$

The three prices per pound can be written in a vector, $\langle 2, 3, 5 \rangle$, and the first equation becomes the dot product:

$$\langle 2, 3, 5 \rangle \cdot \langle x, y, z \rangle = 14.$$

Finally, we have a *matrix equation*:

$$\begin{bmatrix} 1 & 1 & 1 \\ 2 & 3 & 5 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 6 \\ 14 \end{bmatrix}.$$

Without harm, we can make the matrix square:

$$\begin{bmatrix} 1 & 1 & 1 \\ 2 & 3 & 5 \\ 9 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 6 \\ 14 \\ 0 \end{bmatrix}.$$

Example 2.1.4: spreadsheet

One can utilize a spreadsheet and other software to this multiplication for matrices of any dimensions. In order to make this work, the vector X has to be “transposed” (bottom):

		F		•	X	=	Y	
		1	2	3		1		22
		2	-1	4	•	3	=	19
		2	3	5		5		36
		1	0	2				11
				X ^T	=	1	3	5

This is the code for the transpose of X :

```
=TRANSPOSE(R[-5]C:R[-3]C)
```


This is the code for Y :

```
=SUMPRODUCT(RC2:RC4,R8C[-2]:R8C)
```

The whole system can be written in the form of exactly the same matrix equation:

$$FX = B.$$

The multiplication is executed in the same way too:

$$\begin{bmatrix} \text{r} & \text{r} & \text{o} & \text{o} & \text{w} & \text{w} \end{bmatrix} \cdot \begin{bmatrix} \text{c} \\ \text{o} \\ \text{l} \\ \text{u} \\ \text{m} \\ \text{n} \end{bmatrix} = rc + ro + ol + ou + wm + wn$$

Generally, we face a system with:

- 1. the number of variables m , and
- 2. the number of equations n .

We will have an $n \times m$ matrix:

$$\begin{matrix} & \begin{matrix} 1 & 2 & 3 & \dots & m \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ \vdots \\ n \end{matrix} & \begin{bmatrix} 2 & 0 & 3 & \dots & 2 \\ 0 & 6 & 2 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 3 & 1 & 0 & \dots & 12 \end{bmatrix} \end{matrix}$$

Here, the number at the ij -position is the coefficient of the j th variable in the i th equation.

Recall how an index points at a location within a sequence. Similarly, we use *double* index to point at the correct location within a table.

Definition 2.1.5: entries of matrix

The ij -entry in an $n \times m$ matrix A is the number at the i th row and j th column, denoted by

A_{ij}

for each $i = 1, 2, \dots, m$ and each $j = 1, 2, \dots, n$.

For example, we have for the above matrix:

$$\begin{matrix} i \backslash j & \begin{matrix} 1 & 2 & 3 & \dots & n \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ \vdots \\ m \end{matrix} & \begin{bmatrix} A_{1,1} = 2 & A_{1,2} = 0 & A_{1,3} = 3 & \dots & A_{1,n} = 2 \\ A_{2,1} = 0 & A_{2,2} = 6 & A_{2,3} = 2 & \dots & A_{2,n} = 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ A_{m,1} = 3 & A_{m,2} = 1 & A_{m,3} = 0 & \dots & A_{m,n} = 12 \end{bmatrix} \end{matrix}$$

Warning!

What is the difference between tables of numbers and matrices? The algebraic operations discussed here.

2.2. Transformations of the plane

Transformations of the plane are made up of two real-valued functions of two variables.

Definition 2.2.1: transformation of the plane

A *transformation of the plane* is a function

$$F : \mathbf{R}^2 \rightarrow \mathbf{R}^2 ,$$

given by any pair of functions f, g of two variables:

$$F(x, y) = (u, v) = \left(f(x, y), g(x, y) \right)$$

When appropriate, we can also look at the inputs and outputs as *vectors*:

$$\langle x, y \rangle = \begin{bmatrix} x \\ y \end{bmatrix} \mapsto \langle u, v \rangle = \begin{bmatrix} u \\ v \end{bmatrix} ,$$

instead of points.

Let's review a few examples of such transformations.

Example 2.2.2: two functions of two variables

We need two real-valued functions of two variables. Consider $u = f(x, y) = 2x - 3y$:

$$\begin{aligned} f : \mathbf{R}^2 &\rightarrow \mathbf{R}, \text{ meaning} \\ f : (x, y) &\rightarrow u = 2x - 3y \end{aligned}$$

Consider also $v = g(x, y) = x + 5y$:

$$\begin{aligned} g : \mathbf{R}^2 &\rightarrow \mathbf{R}, \text{ meaning} \\ g : (x, y) &\rightarrow v = x + 5y \end{aligned}$$

Let's build a new function from these two. We take the input to be the same – a point in the plane – and we *combine* the two outputs into a single point (u, v) – in another plane. Then what we have is a single function:

$$\begin{aligned} F : \mathbf{R}^2 &\rightarrow \mathbf{R}^2, \text{ meaning} \\ F : (x, y) &\rightarrow (u, v) = (2x - 3y, x + 5y) \end{aligned}$$

In short, this is the formula for this function:

$$F(x, y) = (2x - 3y, x + 5y) .$$

In terms of vectors:

$$F : \begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} 2x - 3y \\ x + 5y \end{bmatrix}$$

The coefficients of the matrix of F are read from that representation:

$$F = \begin{bmatrix} 2 & -3 \\ 1 & 5 \end{bmatrix}$$

What this function does to the plane remains to be determined.

Example 2.2.3: vertical shift

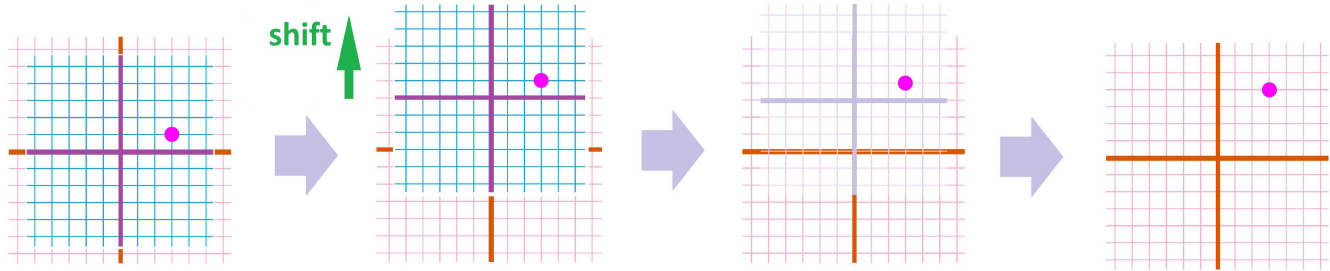
The function defined by

$$F(x,y) = (x,y+3)$$

is a *vertical shift*:

$$(x,y) \xrightarrow{\text{up } k} (x,y+k)$$

We visualize these transformations by drawing something on the original plane (the domain) and then see what that looks like in the new plane (the co-domain):



Predictably, the formula:

$$F(x,y) = (x+a,y+b) = (x,y)+\langle a,b\rangle,$$

gives the *shift by vector* $\langle a,b\rangle$.

Exercise 2.2.4

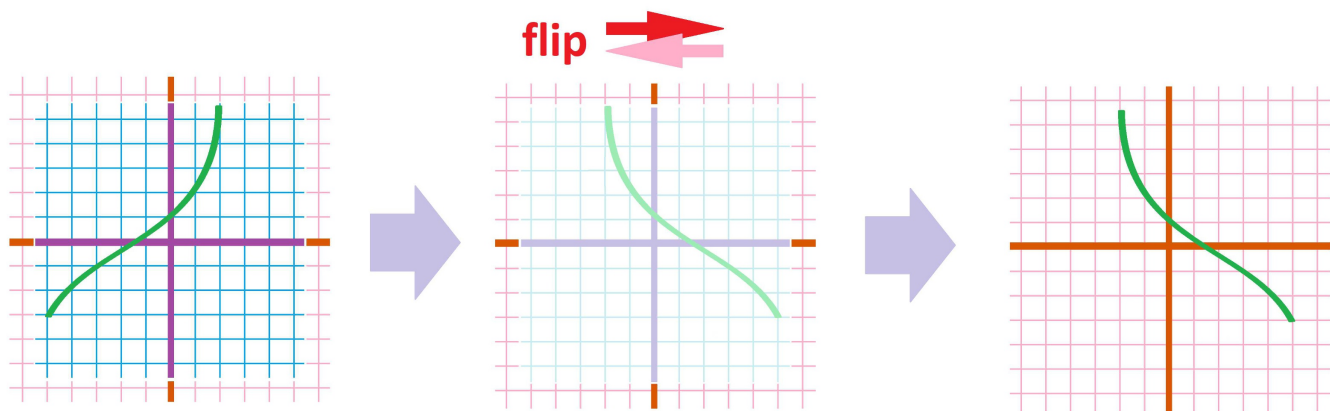
Explain why there is no matrix.

Example 2.2.5: horizontal and vertical flip

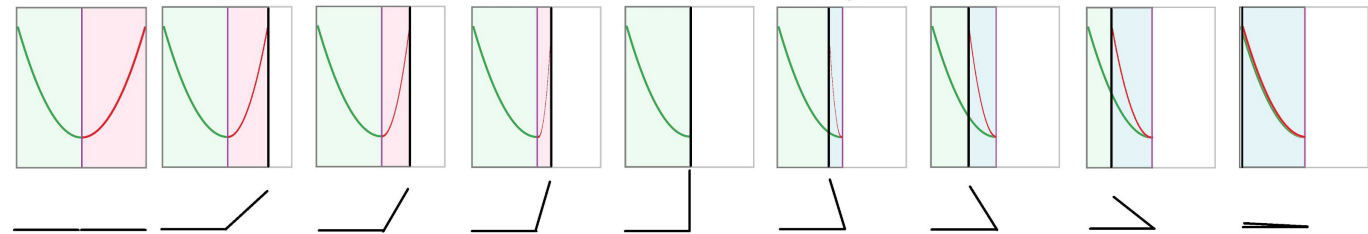
Now the *horizontal flip*. We lift, then flip the sheet of paper with xy -plane on it, and finally place it on top of another such sheet so that the y -axes align. If the function is given by

$$F(x,y) = (-x,y),$$

then we have the following:



Below we illustrate the fact that the parabola's left branch is a mirror image of its right branch:



We can also represent this transformation via vectors:

$$F:\begin{bmatrix}x\\y\end{bmatrix}\mapsto\begin{bmatrix}-x\\y\end{bmatrix}=\begin{bmatrix}(-1)x+0y\\0x+1y\end{bmatrix}$$

Then, we have its matrix:

$$F = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} .$$

Indeed,

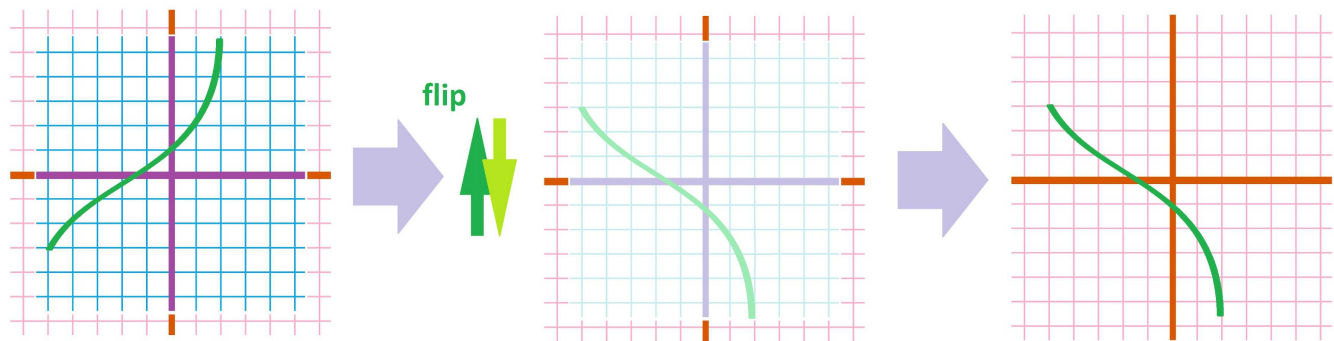
$$F \left(\begin{bmatrix} x \\ y \end{bmatrix} \right) = F \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -x \\ y \end{bmatrix} .$$

Next, consider *vertical flip*. We lift, then flip the sheet of paper with xy -plane on it, and finally place it on top of another such sheet so that the x -axes align. If

$$G(x,y) = (x,-y) ,$$

then we have:

$$(x,y) \xrightarrow{\text{vertical flip}} (x,-y).$$



We can also represent this transformation via a matrix:

$$G = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} .$$

Indeed,

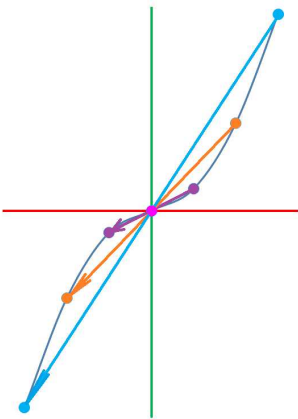
$$G \left(\begin{bmatrix} x \\ y \end{bmatrix} \right) = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ -y \end{bmatrix} .$$

Example 2.2.6: central symmetry

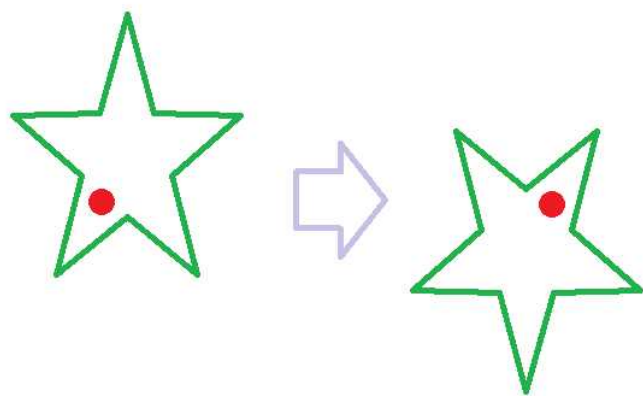
How about the *flip about the origin*? This is the formula,

$$F(x,y) = (-x,-y) ,$$

of what is also known as the central symmetry:



This is what the transformation does to a star:



We can also represent this transformation via vectors:

$$F : \begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} -x \\ y \end{bmatrix} = \begin{bmatrix} (-1)x + 0y \\ 0x + (-1)y \end{bmatrix}$$

Then, we have its matrix:

$$F = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} .$$

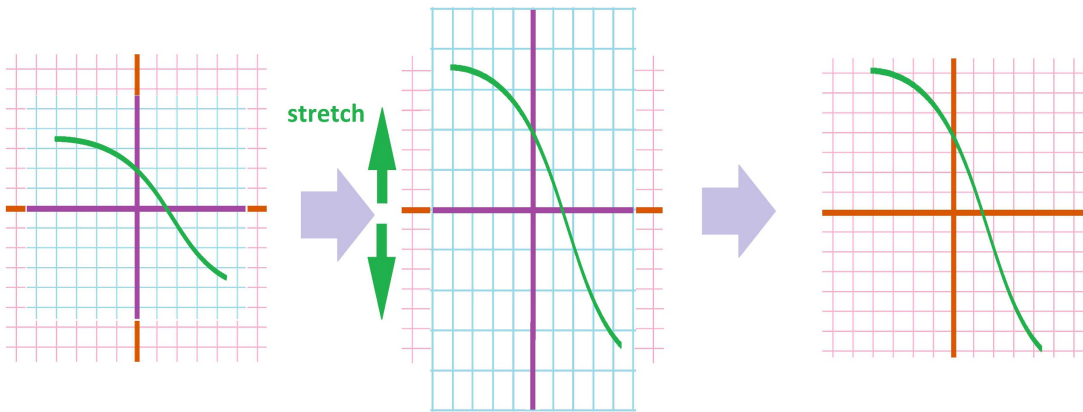
Indeed,

$$F \left(\begin{bmatrix} x \\ y \end{bmatrix} \right) = F \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -x \\ -y \end{bmatrix} .$$

Example 2.2.7: horizontal and vertical stretch

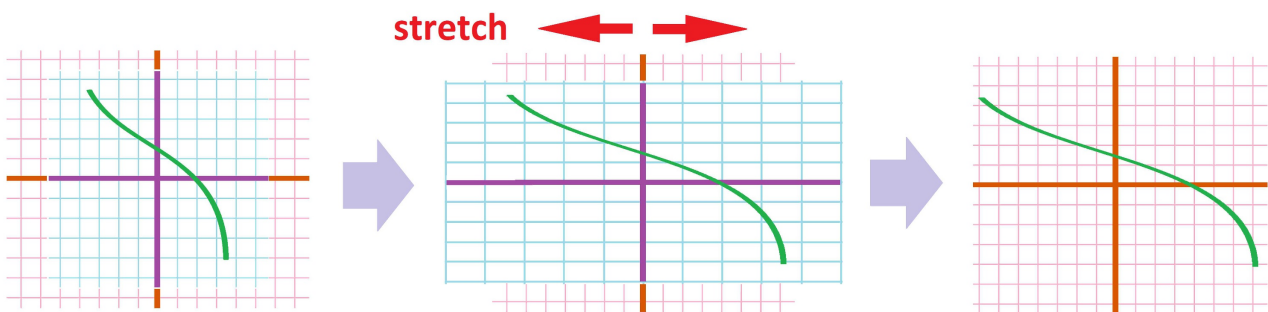
Now the *horizontal stretch*. We grab a rubber sheet by the top and the bottom and pull them apart in such a way that the y -axis doesn't move. Here,

$$F(x,y) = (kx,y) .$$



Similarly, the *horizontal stretch* is given by:

$$G(x,y) = (x,ky) .$$



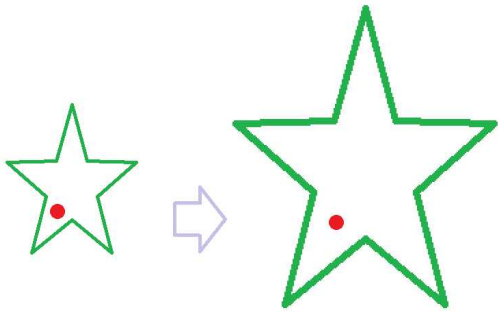
We can also represent these transformations via matrices:

$$F = \begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad G = \begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix} .$$

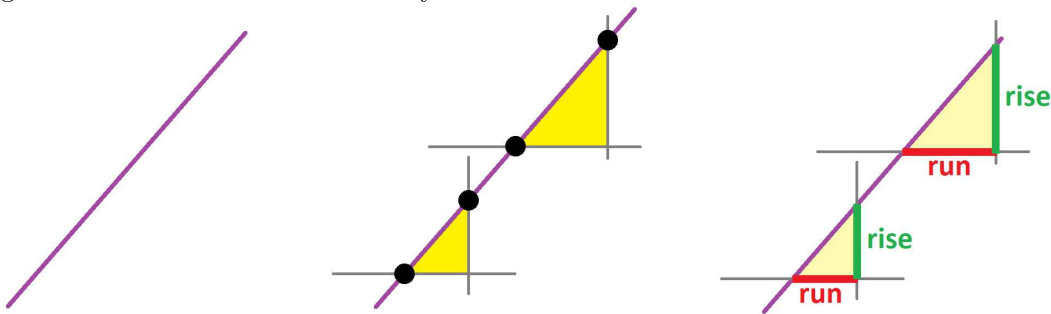
Example 2.2.8: re-scaling

How about the *uniform stretch* (same in all directions)? This is its formula:

$$F(x, y) = (kx, ky) .$$



The result is *re-scaling*. The reason comes from what we know from geometry: Similar triangles have the same angles. Here is an illustration why:



We can also represent this transformations via a matrix:

$$F = \begin{bmatrix} k & 0 \\ 0 & k \end{bmatrix} .$$

Exercise 2.2.9

Find the matrix for a disproportional stretch.

Just as before, we put all newly introduced functions in broad categories.

Some of these categories – such as monotonicity – have become irrelevant.

Others – such as symmetry – have become by far more complex.

Two that will be pursued are one-to-one and onto.

Recall:

- We call a function *one-to-one* if there is no more that one input for each output.
- We call a function *onto* if there is at least one input for each output.

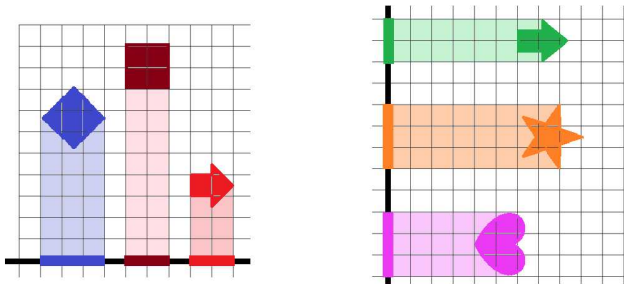
The functions above are all both one-to-one and onto.

Exercise 2.2.10

Prove the last statement.

Example 2.2.11: projections

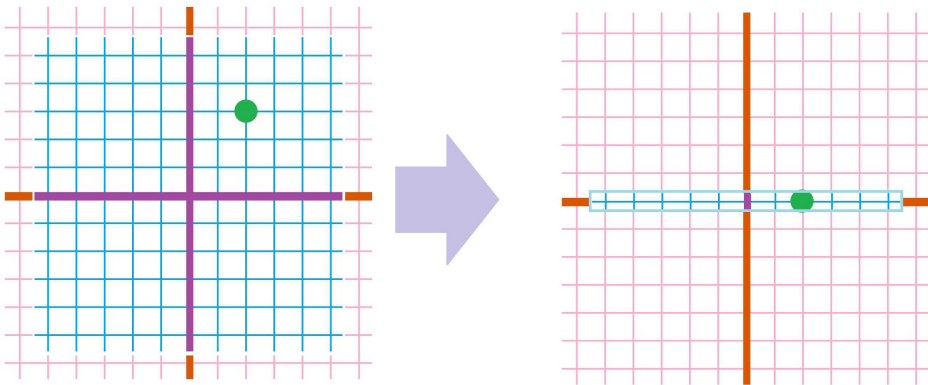
The functions that are not one-to-one or onto are the *projections*. There are at least two types:



This is the vertical one:

$$F(x,y) = (x,0).$$

It is the projection on the x -axis. It's as if the sheet of the xy -plane is rolled into a thin scroll:



We can also represent this transformations via a matrix:

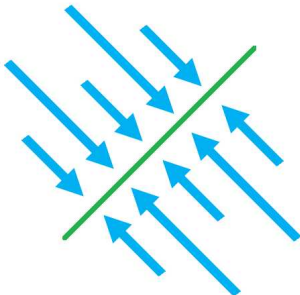
$$F = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Exercise 2.2.12

Find the formula and the matrix for the projection on the y -axis.

Exercise 2.2.13

Find the formula and the matrix for the projection on the line $y = x$.



Example 2.2.14: collapse

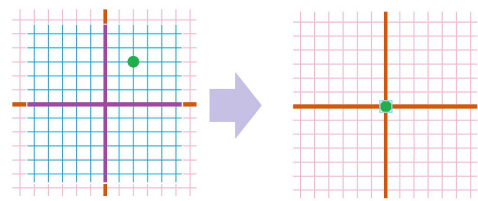
Finally, we have the *collapse*:



It is a constant function:

$$F(x,y)=(x_0,y_0) \, .$$

It’s as if the sheet is crushed into a tiny ball:



There is a matrix representation when this point is the origin:

$$F=\begin{bmatrix}0&0\\0&0\end{bmatrix} \, .$$

Exercise 2.2.15

Show that there is no matrix unless $x_0=0, \, y_0=0$.

The meaning of each number in the matrix depends on its location:

	x	y
x	a	b
y	c	d

This is a special case that we have learned about so far:

Matrix Deconstruction

stretched or flipped x

$\rightarrow a, \, 0$

\leftarrow there is no interaction between x, y

there is no interaction between x, y

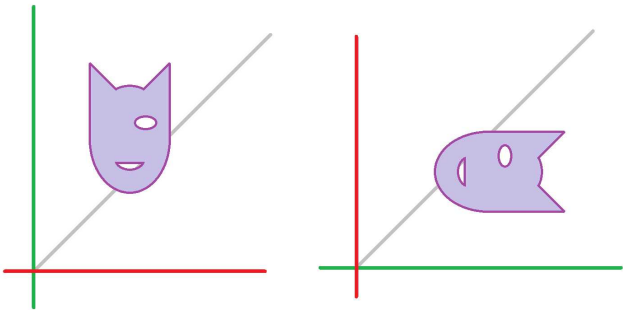
$\rightarrow 0, \, d$

\leftarrow stretched or flipped y

There are many more transformations, however, that aren’t covered so far because they cannot be represented in terms of the vertical and horizontal transformations.

Example 2.2.16: flip about diagonal

A *flip about the line $x = y$* that appeared in the context of finding the graph of the inverse function:

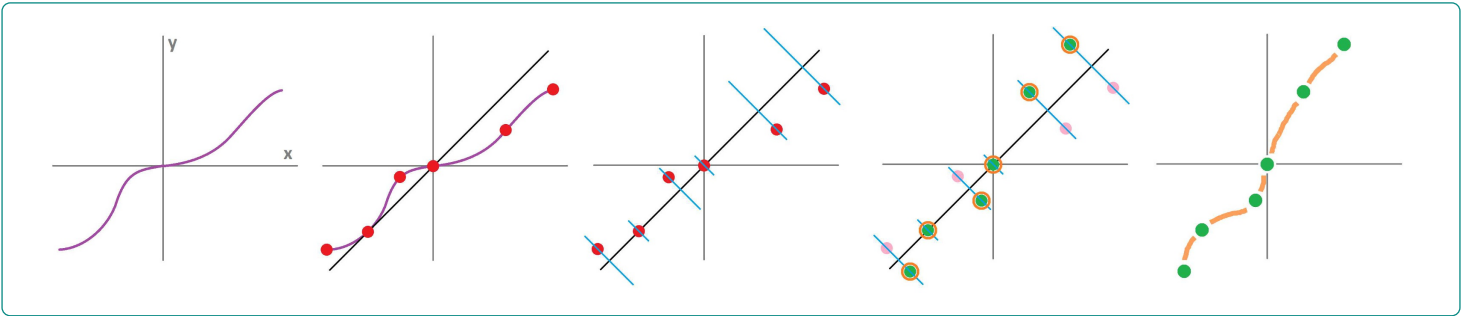


As we acquire the inverse by interchanging x and y , we have the same here:

$$(x,y)\mapsto (y,x) \, .$$

The matrix is:

$$F=\begin{bmatrix}0&1\\1&0\end{bmatrix} \, .$$



We see here some new features:

Matrix Deconstruction

$$\begin{array}{lcl} x \text{ doesn't depends on } x & \rightarrow & 0, \quad 1 \leftarrow x \text{ depends on } y \\ y \text{ depends on } x & \rightarrow & 1, \quad 0 \leftarrow y \text{ doesn't depends on } y \end{array}$$

Exercise 2.2.17

Find the matrix for this rotation:

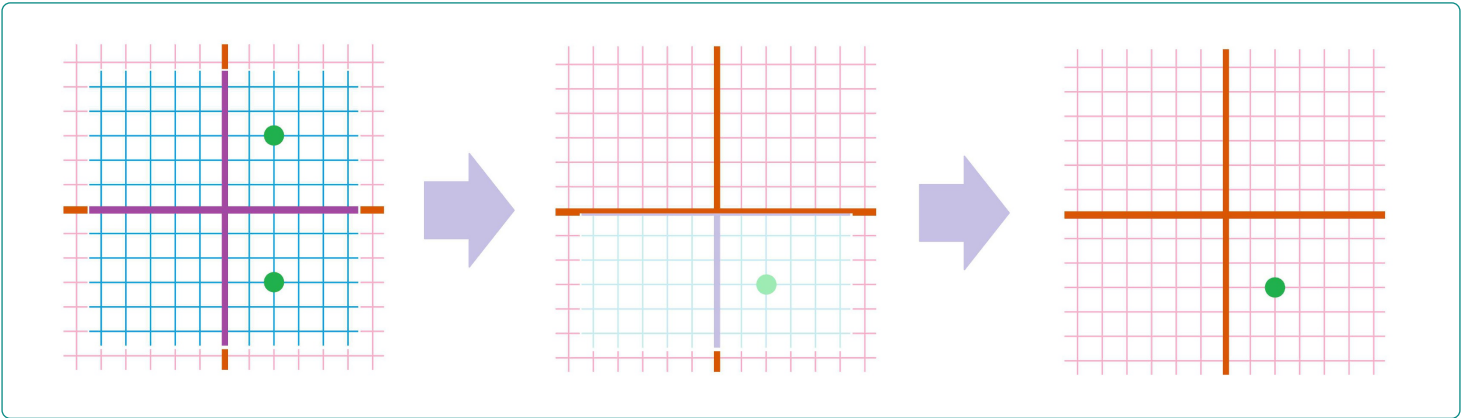
Example 2.2.18: compositions

More complex transformations, however, will require further study. Below, we visualize the 90-degree rotation with a stretch with a factor of 2:

Example 2.2.19: folding

Among others, we may consider *folding* the plane:

$$F(x,y) = (|x|,y) \, .$$



Exercise 2.2.20

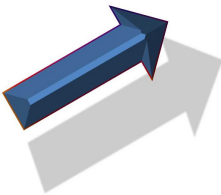
Confirm that this function cannot be represented by a matrix.

Of course, any Euclidean space \mathbf{R}^n can be – in a similar manner – rotated (around various axes), stretched (in various directions), projected (onto various lines or planes), or collapsed.

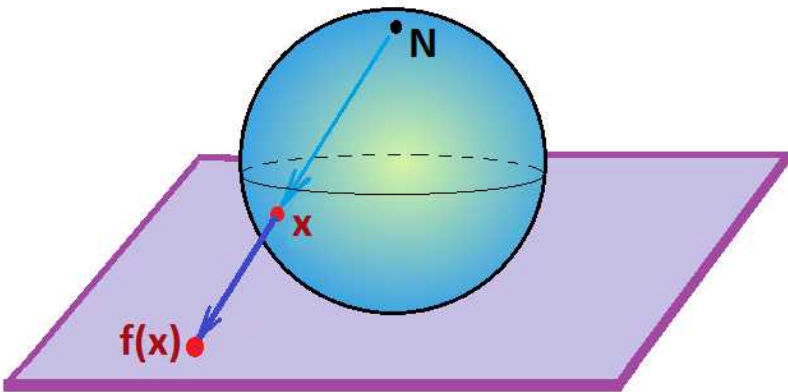
For example, this is how the projection on the xy -plane

$$F(x,y,z) = (x,y,0)$$

works:



This is how maps are made (stereographic projection):



2.3. Linear operators

We consider transformations of the plane as before but we think of the points on the plane as the ends of *2-vectors*. It makes no difference when they are expressed in terms of their components:

$$< x,y > \mapsto < u,v >$$

Then, what happens to the functions? It used to be the case that the coordinates of the output depend on the coordinates of the input:

$$F(x,y) = (u,v) = (2x - 3y, x + 5y).$$

Now our function shows how the components of the output depend on the components of the input:

$$F\big(\langle x,y\rangle\big)=\langle u,v\rangle=\langle 2x-3y,\,x+5y\rangle\,.$$

Here $\langle x,y\rangle$ is the input and $\langle u,v\rangle$ is the output:

input

$\langle x,y\rangle$

\rightarrow

function

F

\rightarrow

output

$\langle u,v\rangle$

Both are vectors; it's a *vector function*:

$$F:\mathbf{R}^2\rightarrow\mathbf{R}^2\,.$$

Warning!

We could interpret F as a vector field, but we won't.

Example 2.3.1: re-scaling

Some of the benefits of this point of view are immediate. For example, even if we didn't know that the transformation given by

$$F(x,y)=(2x,2y)$$

is a uniform stretch, we can discover that fact with our knowledge of vector algebra. We write this function in terms of vectors and discover scalar multiplication:

$$F(\langle x,y\rangle)=\langle 2x,2y\rangle=2\langle x,y\rangle\,.$$

In fact, this idea will work in any dimension. This is how you stretch the space by a factor of 2, using the component-free approach:

$$F:\mathbf{R}^n\rightarrow\mathbf{R}^n\text{ given by }F(X)=2X\,.$$

The approach gives us a better representation when the functions that make up the transformation happen to be *linear*. Matrices rely on vector algebra:

$$F(X)=(f(x,y),g(x,y))=\begin{bmatrix}a&b\\c&d\end{bmatrix}\cdot\begin{bmatrix}x\\y\end{bmatrix}=F\cdot X\,.$$

Exercise 2.3.2

Does it work when the function isn't linear? Try $F(x,y)=\left(e^x,\frac{1}{y}\right)$.

Conversely, if we do have a matrix, we can always understand it as a function, as follows:

$$\begin{bmatrix}a&b\\c&d\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}ax+by\\cx+dy\end{bmatrix}=\langle ax+by,\,cx+dy\rangle\,,$$

for some a,b,c,d fixed. So, matrix F contains all the information about function F . One can think of F (a table) as an abbreviation of FX (a formula).

Warning!

We will continue to use the same letter for the function and the matrix.

Clearly, a function given by a matrix is a *special* one. What is so special about it?

Now, the domain \mathbf{R}^2 of this function is a *vector space*, and so is its codomain. How does such a function interact with the algebra of these two spaces? What happens to the *vector operations* under F ?

Suppose we have addition and scalar multiplication carried out in the domain space of F . Once F has transformed the plane, what do these operations look like now?

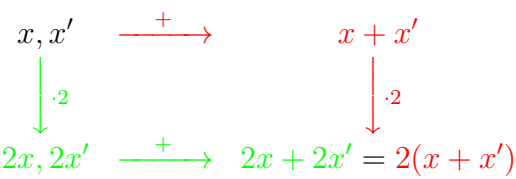
Example 2.3.3: dimension 1

The multiplication by 2, $f(x) = 2x$ “preserves addition”:

$$f(x + x') = 2(x + x') = 2x + 2x' = f(x) + f(x') .$$

After all, this is just a stretch by a factor of 2.

The computation is just an abbreviation of the following diagram:



In the diagram, we start with a pair of numbers at the top left and then we proceed in two ways:

- Right: Add them. Then down: Apply the function to the result.
- Down: Apply the function to them. Then right: Add the results.

A shift by 1, $f(x) = x + 1$, doesn’t preserve addition:

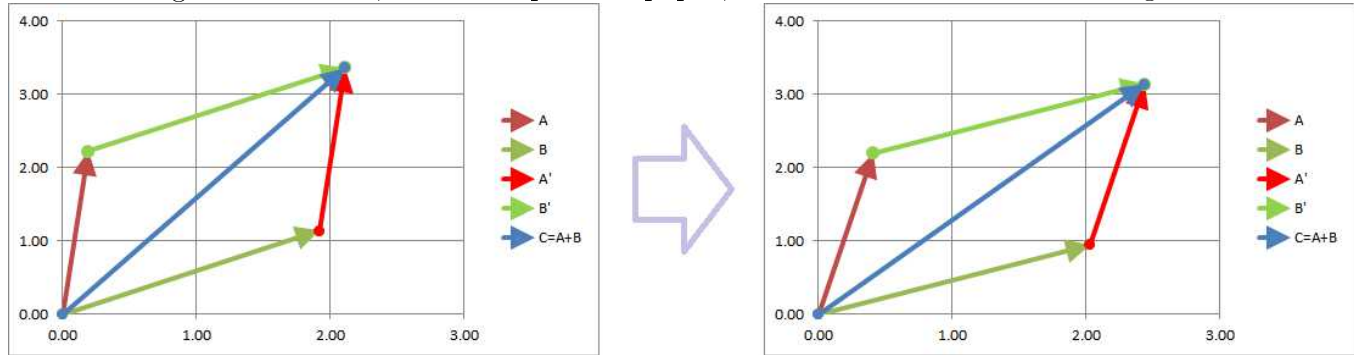
$$f(x + x') = (x + x') + 1 = x + x' + 1 \neq x + x' + 2 = (x + 1) + (x' + 1) = f(x) + f(x') .$$

Exercise 2.3.4

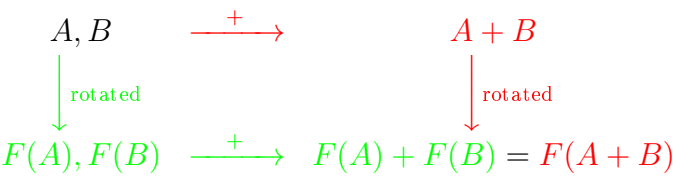
What effect does the function $f : \mathbf{R} \rightarrow \mathbf{R}$ given by $f(x) = 2x$ have on multiplication in its domain?

Example 2.3.5: addition under rotation

What happens to an addition diagram (the parallelogram construction) when the *plane* is transformed? If such a diagram is *rotated*, as if on a piece of paper, it will remain an addition diagram:



We see the parallelogram rule of addition on the left and on the right. This is the algebra:



What happens to an addition diagram when the vector space is transformed? When it is still an addition diagram, this is the language we will use:

Definition 2.3.6: addition is preserved

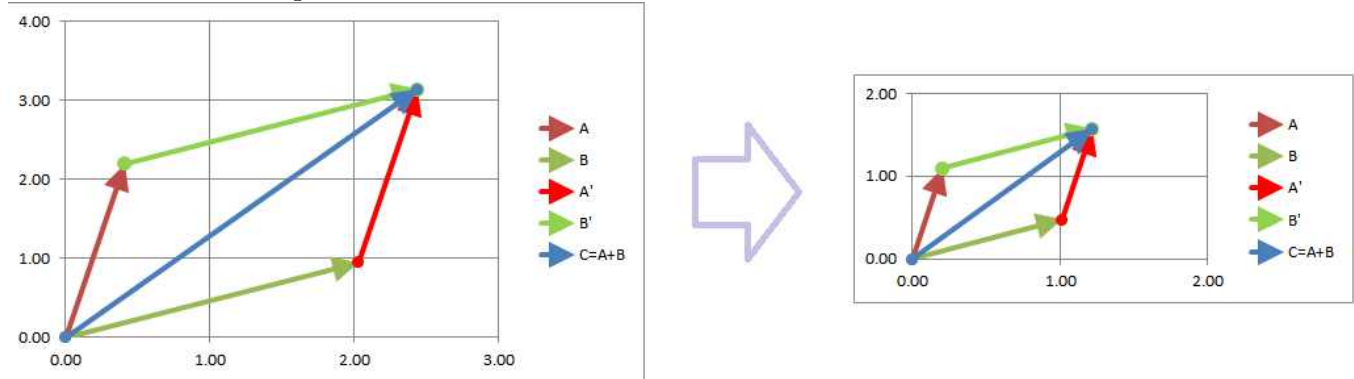
We say that *addition is preserved* under a function $F : \mathbf{R}^n \rightarrow \mathbf{R}^m$ if

$$F(A + B) = F(A) + F(B)$$

for any vectors A and B .

Example 2.3.7: stretch dimension 2

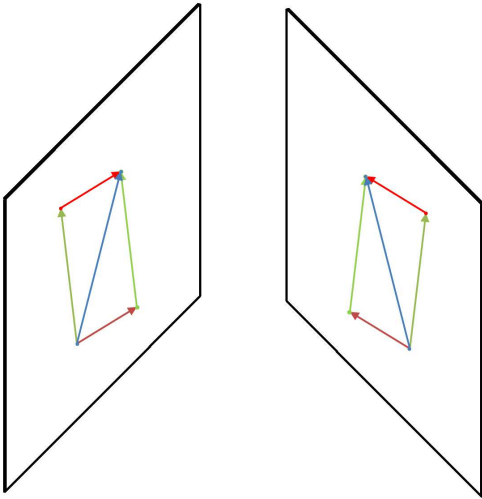
Furthermore, the example of a stretch below shows that the triangles of the diagram, if not identical, are *similar* to the original:



It's as if we just stepped away from the piece of paper that has the addition diagram on it or put the diagram under a magnifying glass.

Example 2.3.8: reflection dimension 2

We can also see the addition diagram in the mirror, and it's still an addition diagram:



Exercise 2.3.9

Show that a fold doesn't preserve vector addition. Suggest other examples.

What about the general case?

Theorem 2.3.10: Preserving Addition

If a function $F : \mathbf{R}^n \rightarrow \mathbf{R}^m$ is given by a matrix, $F(X) = FX$, it preserves addition.

Proof.

Consider $F : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ and two input vectors:

$$A = \begin{bmatrix} x \\ y \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} x' \\ y' \end{bmatrix} .$$

Let’s confirm the formula:

$$F(A + B) = F(A) + F(B) .$$

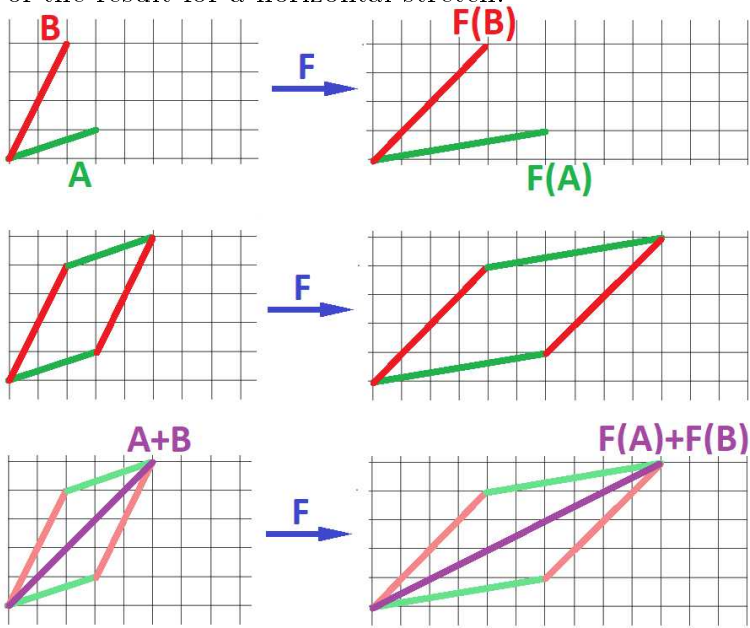
Let’s compare:

Left-hand side:	Right-hand side:
$F \left(\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} x' \\ y' \end{bmatrix} \right)$	$F \begin{bmatrix} x \\ y \end{bmatrix} + F \begin{bmatrix} x' \\ y' \end{bmatrix}$
$= \begin{bmatrix} a & b \\ c & d \end{bmatrix} \left(\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} x' \\ y' \end{bmatrix} \right)$	$= \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix}$
$= \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x + x' \\ y + y' \end{bmatrix}$	$= \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix} + \begin{bmatrix} ax' + by' \\ cx' + dy' \end{bmatrix}$
$= \begin{bmatrix} a(x + x') + b(y + y') \\ c(x + x') + d(y + y') \end{bmatrix}$	$= \begin{bmatrix} ax + by + ax' + by' \\ cx + dy + cx' + dy' \end{bmatrix}$

These are the same, after factoring.

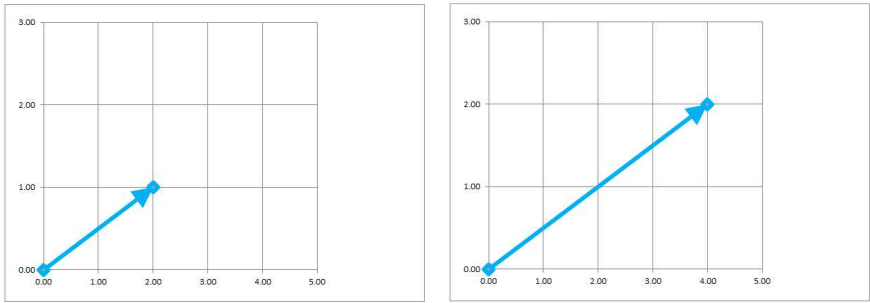
Example 2.3.11: stretch dimension 2

Here is a confirmation of the result for a horizontal stretch:



We simply compare what happens in the domain with its “reflection” in the codomain.

The diagram of scalar multiplication is much simpler:



It is a stretch of the vector; rotated or stretched, it remains a stretch. This is the language we will use:

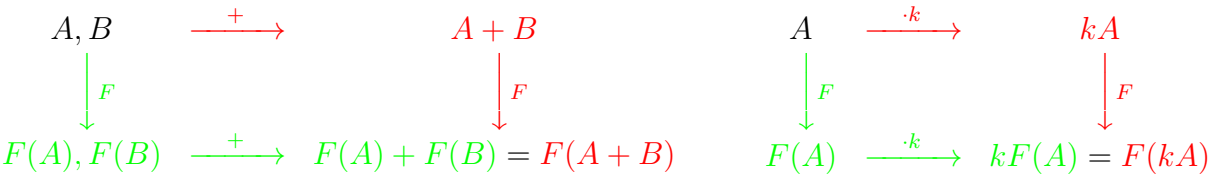
Definition 2.3.12: scalar multiplication is preserved

We say that *scalar multiplication is preserved* under a function $F : \mathbf{R}^n \rightarrow \mathbf{R}^m$ if

$$F(kA) = kF(A)$$

for any vector X and real k .

The formulas in the two definitions are just abbreviations of these diagrams:



In other words, the order of these operations makes no difference.

Example 2.3.13: motions

The same conclusion is quickly reached for the flip and other motions: The triangles of the new diagram are identical to the original. We can just imagine that the addition diagram is drawn on a piece of paper with no grid, which then has been rotated:

Theorem 2.3.14: Preserving Scalar Multiplication

If a function $F : \mathbf{R}^n \rightarrow \mathbf{R}^n$ is given by a matrix, $F(A) = FA$, it preserves scalar multiplication.

Proof.

Consider an input vector $X = \langle x, y \rangle$ and a scalar k . Then,

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \left(k \begin{bmatrix} x \\ y \end{bmatrix} \right) = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} kx \\ ky \end{bmatrix} = \begin{bmatrix} akx + bky \\ ckx + dky \end{bmatrix} = k \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix} = k \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} .$$

Now, this is the general case:

Definition 2.3.15: linear operator

A function $F : \mathbf{R}^n \rightarrow \mathbf{R}^m$ that preserves both addition and scalar multiplication is called a *linear operator* (or a linear map); i.e.,:

$$\begin{aligned} F(U + V) &= F(U) + F(V) \\ F(kV) &= kF(V) \end{aligned}$$

Warning!

Previously, $y = ax + b$ has been called a “linear function”. Now, $y = ax$ is called a “linear operator”.

We combine the two operations together:

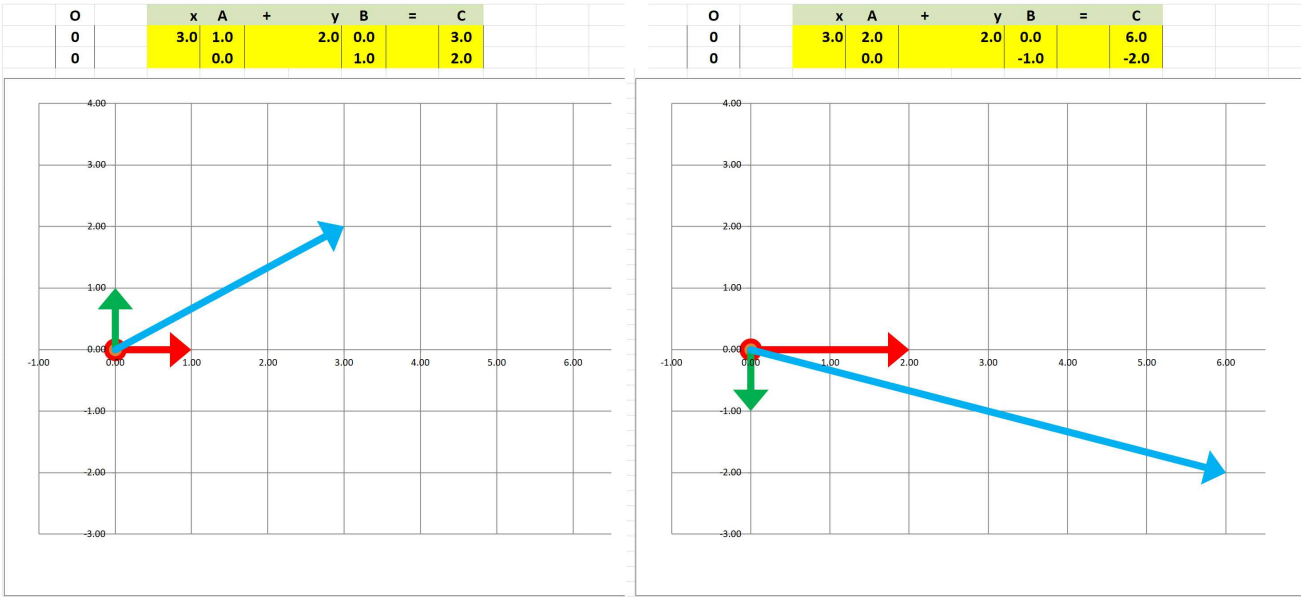
Theorem 2.3.16: Linear Operators and Linear Combinations

A linear operator $F : \mathbf{R}^n \rightarrow \mathbf{R}^m$ preserves linear combinations; i.e.,

$$X = xA + yB \implies F(X) = xF(A) + yF(B),$$

for any vectors A and B and any real coefficients x and y .

In other words, the diagram of a linear combination will remain such under a linear operator:



Exercise 2.3.17

Describe what this linear operator does and find its matrix.

This is the summary of our analysis.

Theorem 2.3.18: Linear Operators vs. Matrices

- The function $F : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ defined via multiplication by a 2×2 matrix F ,
$$F(X) = FX,$$
is a linear operator.

- Conversely, every linear operator $F : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ is defined via multiplication by some 2×2 matrix.

Proof.

The first part of the theorem follows from the two theorems above. The converse is proven in the next section.

Warning!

Linear operators and matrices aren't interchangeable, because matrices emerge only when a Cartesian system has been specified.

Corollary 2.3.19: Linear Operator at 0

A linear operator takes the zero vector to the zero vector:

$$F(0) = 0.$$

Exercise 2.3.20

Prove the corollary (a) from the definition of a linear operator, (b) by examining the matrix multiplication.

The conclusion will be visible in the examples in the next section.

The latter part of the theorem is materialized when a matrix is found for a linear operator described by what it does. We start with a couple of simple examples.

Let's not forget:

► Linear operators are functions.

The two simplest functions – no matter what the domain and codomain are – are the constant function and the identity function.

However, according to the last corollary, there can be only one constant, 0, and, therefore, only one constant linear operator. This is the simplest linear operator:

Definition 2.3.21: zero operator

The function $F : \mathbf{R}^n \rightarrow \mathbf{R}^m$ that takes every vector to the zero vector,

$$F(0) = 0,$$

is called the *zero operator*. The notation is as follows:

$$0 : \mathbf{R}^n \rightarrow \mathbf{R}^m, \quad 0(X) = 0.$$

The matrix of the zero operator consists, of course, of all zeros:

$$0 = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}$$

We can write:

$$0_{ij} = 0 .$$

What about the identity function? The dimensions of the domain and the codomain must be the same:

Definition 2.3.22: identity operator

The function $F : \mathbf{R}^n \rightarrow \mathbf{R}^n$ that takes every vector to itself,

$$F(X) = X ,$$

is called the *identity operator*. The notation is as follows:

$$I : \mathbf{R}^n \rightarrow \mathbf{R}^n, \quad I(X) = X .$$

The matrix of the identity operator is the following:

$$I = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

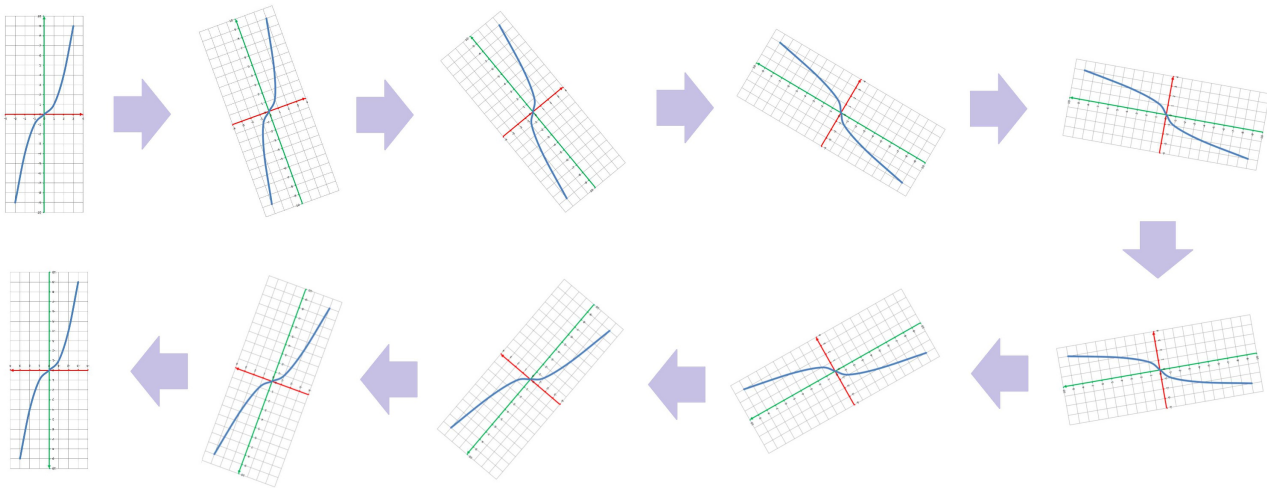
It has 1's on the main diagonal and 0's elsewhere. We can write:

$$I_{ij} = \begin{cases} 1 & \text{if } i = j , \\ 0 & \text{if } i \neq j . \end{cases}$$

2.4. Examining and building linear operators

The transformations of the plane are illustrated previously with curves plotted on the original plane and then seen transformed. Beyond just saying that the curve was drawn on a piece of paper or a sheet of rubber, what exactly happens to those curves?

Using the graphs of functions to represent these curves in the domain of the transformation fails. For example, rotating such a curve, $y = f(x)$, is likely to produce a curve that isn't the graph of any function, $u = g(v)$, in the codomain:



Our choice is, then, *parametric curves*:

$$P : \mathbf{R} \rightarrow \mathbf{R}^2,$$

given by:

$$X = P(t) \text{ or } x = x(t), \, y = y(t) .$$

Example 2.4.1: non-uniform re-scale

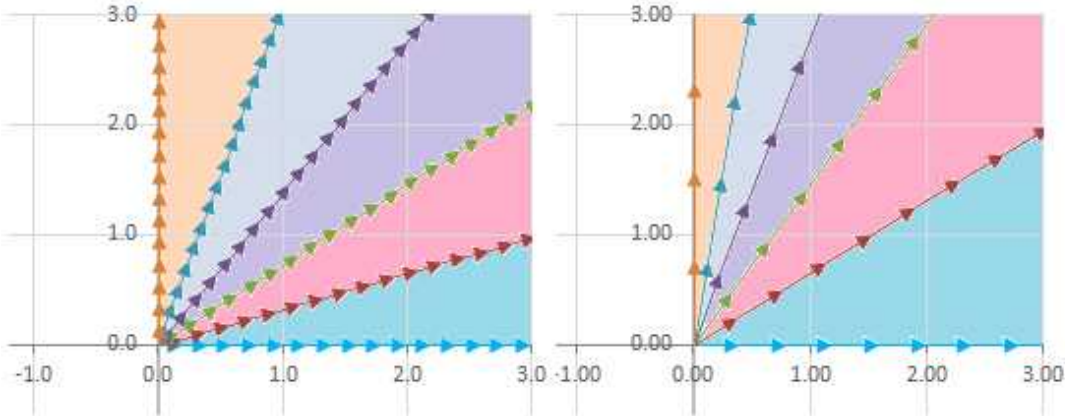
Let's consider this transformation:

$$\begin{cases} u &= 2x \\ v &= 4y \end{cases}$$

Here, this function is given by the matrix:

$$F = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$$

We can see what happens to the lines through 0:

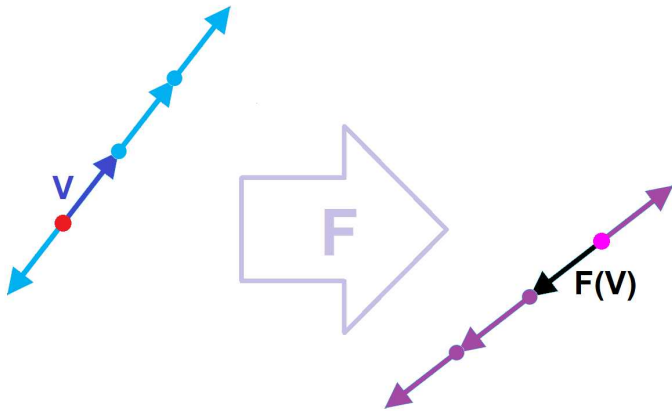


They all remain straight even though some of them rotate.

Let's take a closer look at the straight lines that pass through the origin. The equation of such a line is very simple:

$$P(t) = tV ,$$

where $V \neq 0$ is some fixed vector. As V is transformed by F , then so are all of its multiples:



It's another straight line. Linear operators don't bend!

We confirm the observation:

Theorem 2.4.2: Images of Lines

The image of a straight line through the origin under a linear operator will produce another straight line thorough the origin.

Proof.

Suppose F is such an operator and V is the direction vector of the line. Then:

$$P(t) = tV \xrightarrow{F} (F \circ P)(t) = F(tV) = tF(V)$$

This is a line with $F(V)$ as the direction vector.

In general, we witness the following:

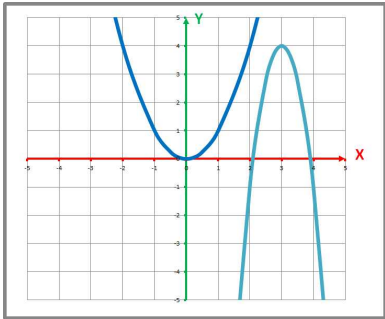
$$\text{parametric curve} \xrightarrow{F} \text{parametric curve}$$

Let’s consider two more basic types of curves.
The graph of a quadratic polynomial is called a “parabola”. This is what we care about:

- Any parabola can be acquired from *the* parabola of $f(x) = x^2$ via a vertical stretch or shrink.

It follows from the fact that every quadratic polynomial has its *vertex form*:

$$f(x) = a(x - h)^2 + k .$$

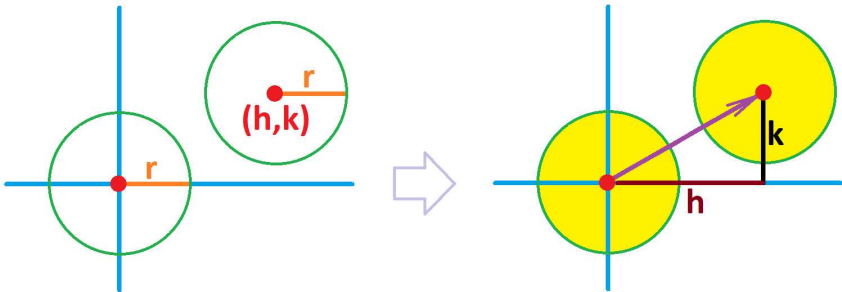


There are many circles on the plane. This is what we care about:

- Any circle can be acquired from *the* circle of $x^2 + y^2 = 1$ via a uniform stretch or shrink.

It follows from the fact that every circle has its *centered form*:

$$(x - h)^2 + (y - k)^2 = r^2 .$$



Exercise 2.4.3

Explain how the graph of any exponential function a^x can be acquired from the natural one e^x via linear transformations. Same for the logarithms.

We take advantage of our knowledge of vector algebra to summarize these three cases:

	“template”	relation	parametric
/	line: $y = x$	$y - 3 = 2(x - 1)$ vertical stretch by 2, shift up by $\langle 1, 3 \rangle$	$(x, y) = (1, 3) + t \langle 1, 2 \rangle$
∪	parabola: $y = x^2$	$y - 3 = 2(x - 1)^2$ vertical stretch by 2, shift up by $\langle 1, 3 \rangle$	$(x, y) = (1, 3) + \langle t, 2t^2 \rangle$
○	circle: $x^2 + y^2 = 1$	$(y - 3)^2 + (x - 1)^2 = 2^2$ uniform stretch by 2, shift up by $\langle 1, 3 \rangle$	$(x, y) = (1, 3) + 2 \langle \cos t, \sin t \rangle$

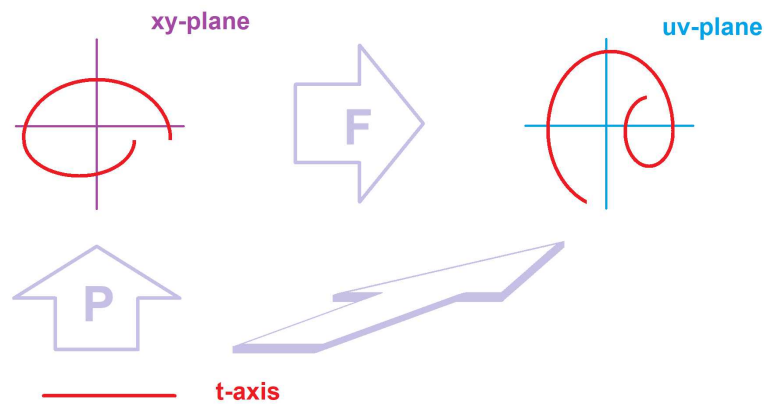
We will need to visualize examples of linear operators on the plane:

$F : \mathbf{R}^2 \rightarrow \mathbf{R}^2 .$

We illustrate them by creating marks on the original plane and then see what happens to them as they appear on the new plane. Each of these marks will be a parametric curve in the domain:

$P : \mathbf{R} \rightarrow \mathbf{R}^2, \; X = P(t) .$

We then plot its image in the codomain under the transformation $Y = F(X)$:



To be precise, we plot the image of the *composition* of the two functions:

$F \circ P : \mathbf{R} \rightarrow \mathbf{R}^2, \; Y = F(P(t)) .$

It is also a parametric curve.

In other words, we have:

$$\begin{array}{ccc} \mathbf{R}^2 & \xrightarrow{F} & \mathbf{R}^2 \\ \uparrow_P & \nearrow_{F \circ P} & \\ \mathbf{R} & & \end{array}$$

We will plot many pairs of curves each time:

$$\begin{array}{ccccc} & \mathbf{R}^2 & & & \mathbf{R}^2 \\ \text{old curve: } & \uparrow_P & \longrightarrow & \text{new curve: } & \nearrow_{F \circ P} \\ & \mathbf{R} & & \mathbf{R} & \end{array}$$

We can use this setup in two ways:

- 1. We can study a curve by applying various transformations to the plane.
- 2. We can study a transformation by applying it to various curves.

We did the former in the beginning of the section. Now the latter.

Example 2.4.4: stretch

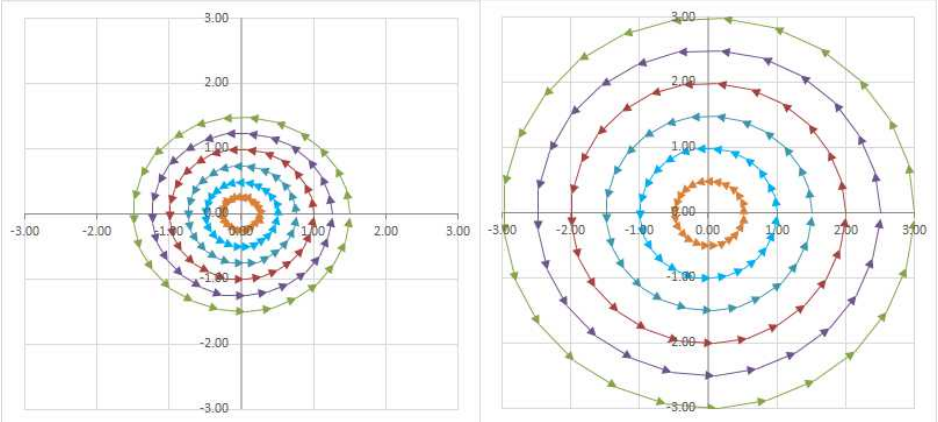
So, applying transformations to curves will give us new curves. For example, we start with the circles:

$$P(t) = r \langle \cos t, \sin t \rangle, \quad r > 0.$$

Then, using scalar multiplication by 2 on all vectors means *stretching radially* the whole space. We then discover that the image of the curve is given by:

$$Q(t) = 2P(t) = 2r \langle \cos t, \sin t \rangle.$$

It is a parametric curve of the circle of radius $2r$:

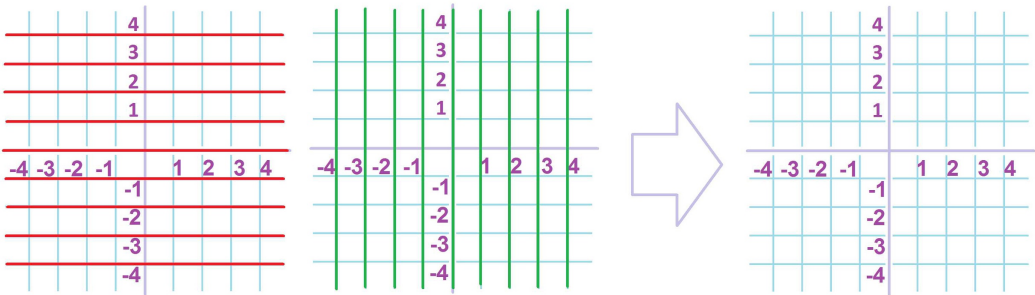


What curves do we choose? Familiar ones: straight lines and circles. How many? The whole *grid*. We have two possibilities:

Cartesian grid: a rectangular grid of lines

Polar grid: a grid of circles and radii

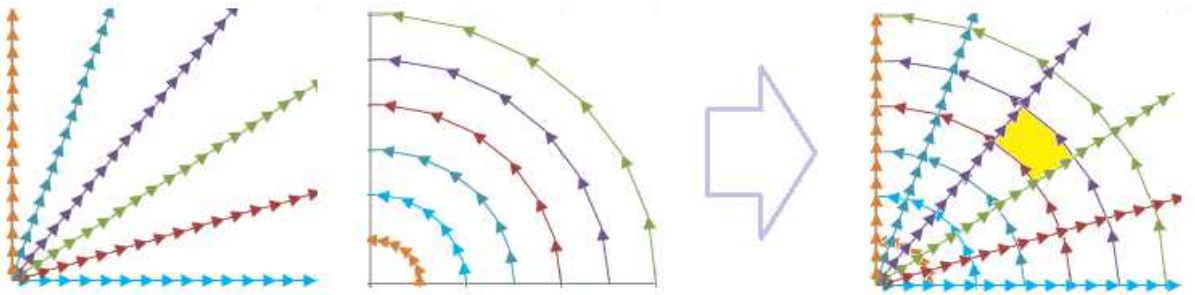
The Cartesian grid is created with these two types of lines:



These lines are defined parametrically:

1. horizontal: $x = t, \quad y = k, \quad k = \dots - 3, -2, -1, 0, 1, 2, 3, \dots$
2. vertical: $x = k, \quad y = t, \quad k = \dots - 3, -2, -1, 0, 1, 2, 3, \dots$

The polar grid is created with these two types of lines:



Here they are defined parametrically:

1. rays: $x = at, \ y = bt, \ a, b$ real; and
2. circles: $x = r \cos t, \ y = r \sin t, \ r$ real.

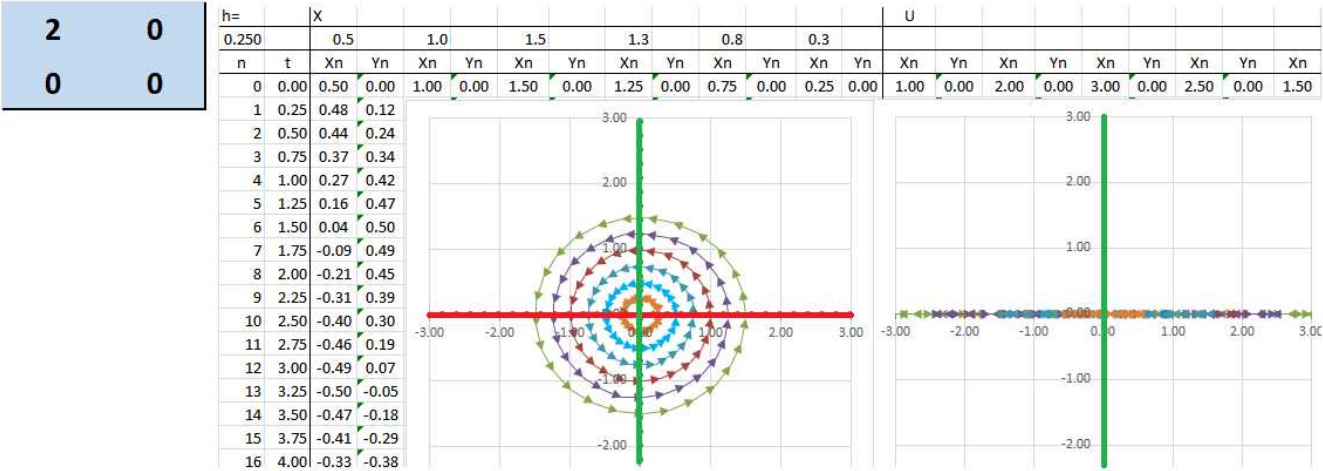
We will apply the Cartesian or the polar as needed.

Example 2.4.5: collapse on axis

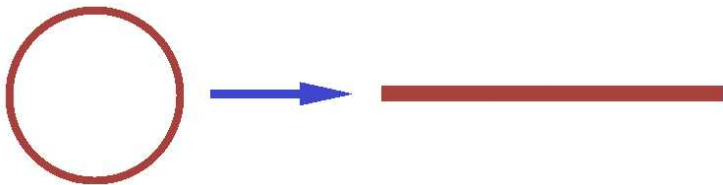
Let’s consider this very simple function:

$$\begin{cases} u &= 2x \\ v &= 0 \end{cases} \text{ , re-written: } \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} .$$

Below, one can see how this function collapses the whole plane to the x -axis:



This is what happens to every circle:



In the meantime, the x -axis is stretched by a factor of 2. We can see both in the matrix:

stretch of $x \rightarrow 2, \ 0 \leftarrow$

y doesn't depends on $x \rightarrow 0, \ 0 \leftarrow$

} collapse

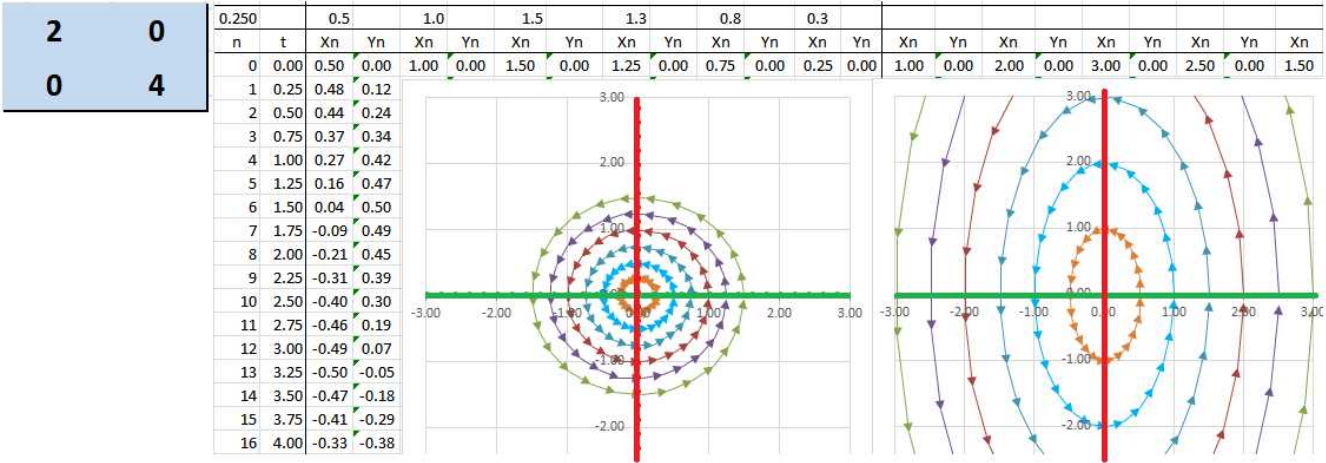
Because of the collapse of the y -axis, the function is neither one-to-one nor onto.

Example 2.4.6: stretch-shrink along axes

Let’s revisit this linear operator:

$$F = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} .$$

We look at the circles first:

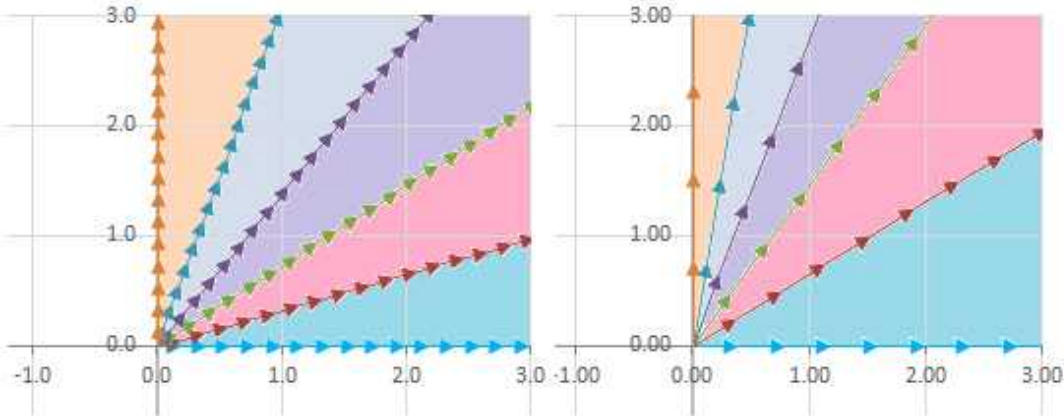


The circles have become ellipses! We can see what happens in the matrix:

stretch of $x \rightarrow 2, 0 \leftarrow x$ doesn't depends on y

y doesn't depends on $x \rightarrow 0, 4 \leftarrow$ stretch of y

The axes stay put. What happens to the rest of the plane? Let's look at the lines now:



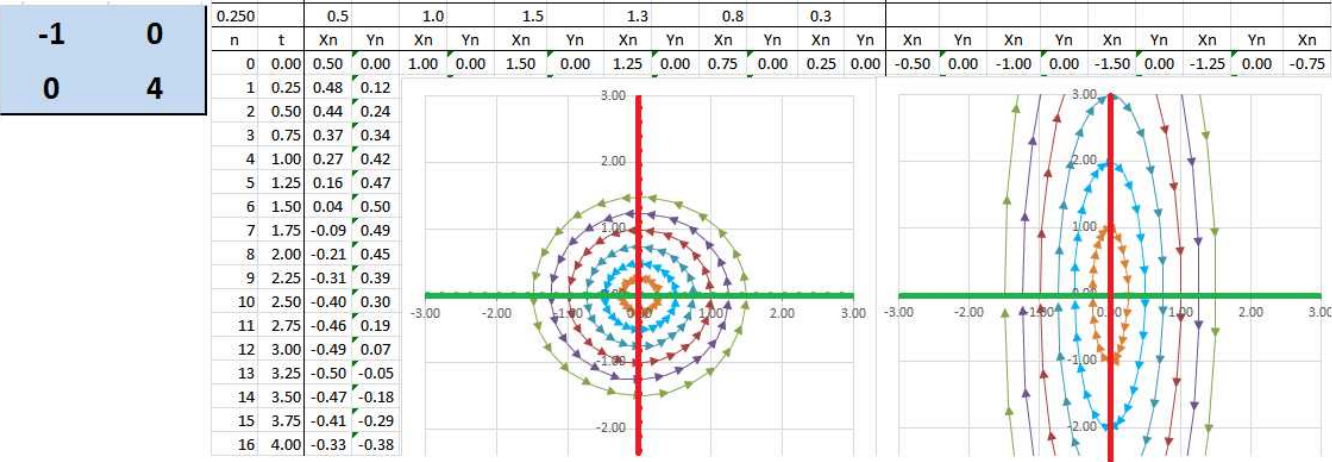
Since the stretching is non-uniform, the vectors turn. However, since the basis vectors e_1 and e_2 don't turn, this is not a rotation but rather a “fanning out” of the vectors. Their slopes have increased. We also discover that the function is both one-to-one and onto.

Example 2.4.7: stretch-shrink along axes

A slightly different function is:

$$\begin{cases} u &= -x \\ v &= 4y \end{cases}$$

It is simple because the two variables are fully separated. Just the circles:



The slight change to the function produces a similar but different pattern: We see the reversal of the direction of the ellipse around the origin. We way that *the orientation has changed*. The matrix of F

is still diagonal:

$$F = \begin{bmatrix} -1 & 0 \\ 0 & 4 \end{bmatrix}.$$

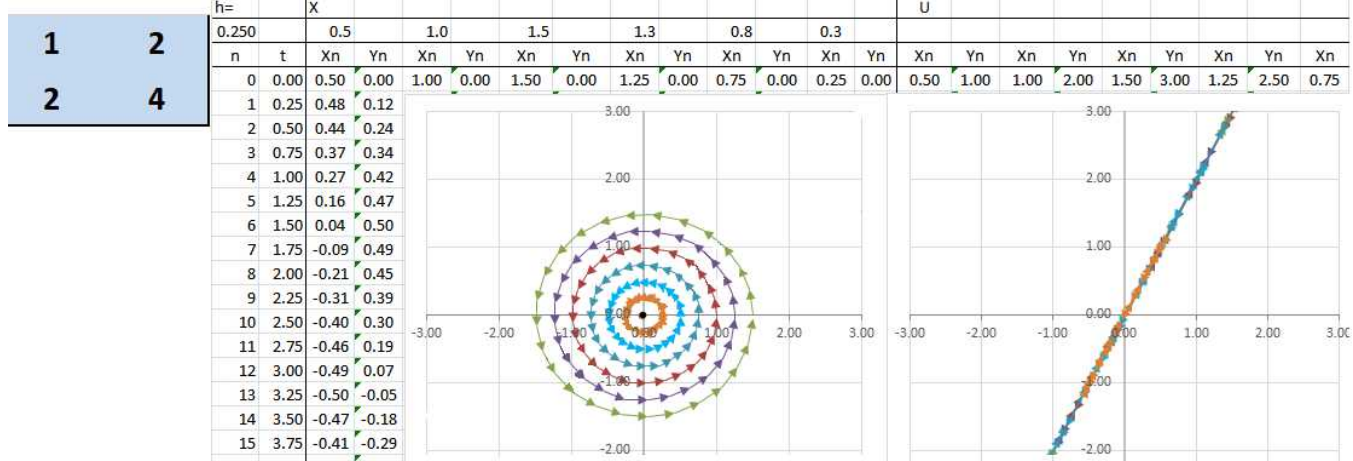
The function is both one-to-one and onto.

Example 2.4.8: experiment

Let’s consider a more general function:

$$\begin{cases} u = x + 2y \\ v = 2x + 4y \end{cases} \implies F = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}.$$

It is hard to tell what it does, judging by its matrix. We *experiment*:



It appears that the function is stretching the plane in one direction and collapsing in another. That’s why there is a whole line of points X with $FX = 0$. To find it, we solve this equation:

$$\begin{cases} x + 2y = 0 \\ 2x + 4y = 0 \end{cases} \implies x = -2y.$$

The vector $\langle 1, 2 \rangle$ is, in fact, visible in the matrix. Because of the collapse of the green line to the origin, the function is neither one-to-one nor onto.

Exercise 2.4.9

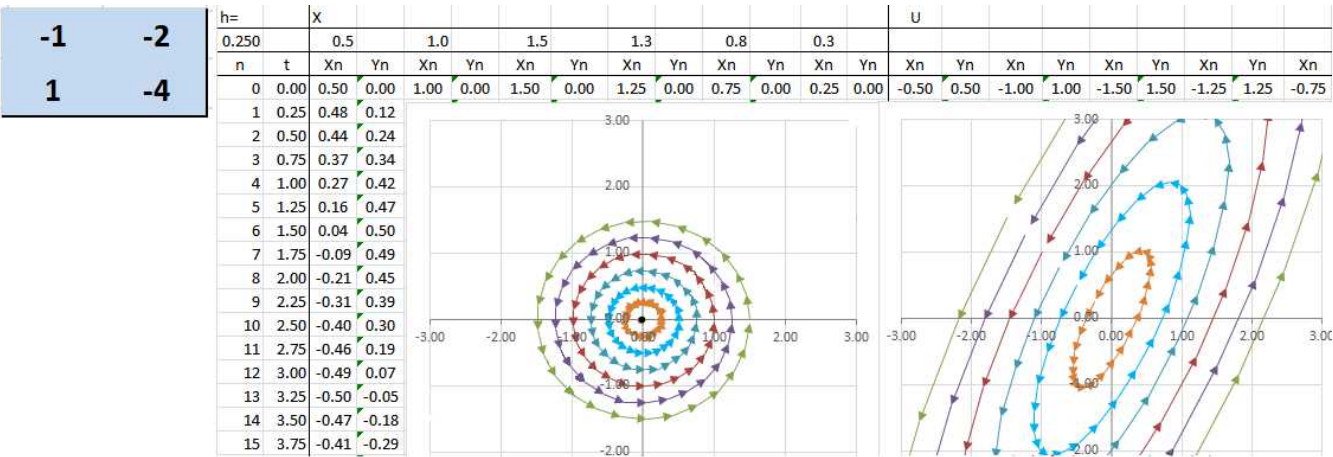
Show where each point goes.

Example 2.4.10: experiment

Consider the following matrix F :

$$F = \begin{bmatrix} -1 & -2 \\ 1 & -4 \end{bmatrix}.$$

Again, it’s too complex to reveal what it does, and we have to experiment:



It looks like a non-uniform stretch along diagonal directions. The function is both one-to-one and onto.

Exercise 2.4.11

Describe what is happening under this operator F :

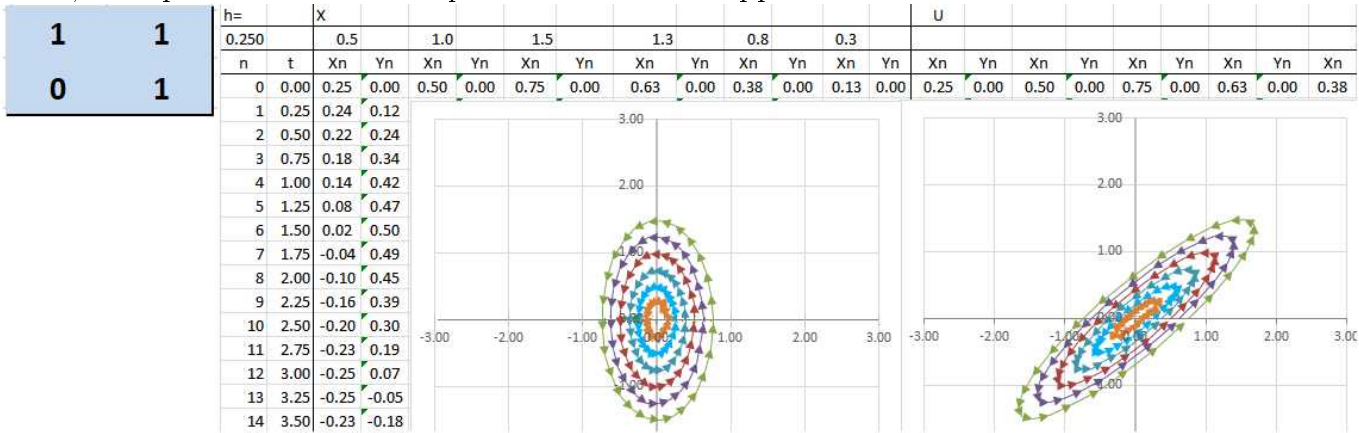
$$F = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}.$$

Example 2.4.12: skewing-shearing

Consider this matrix:

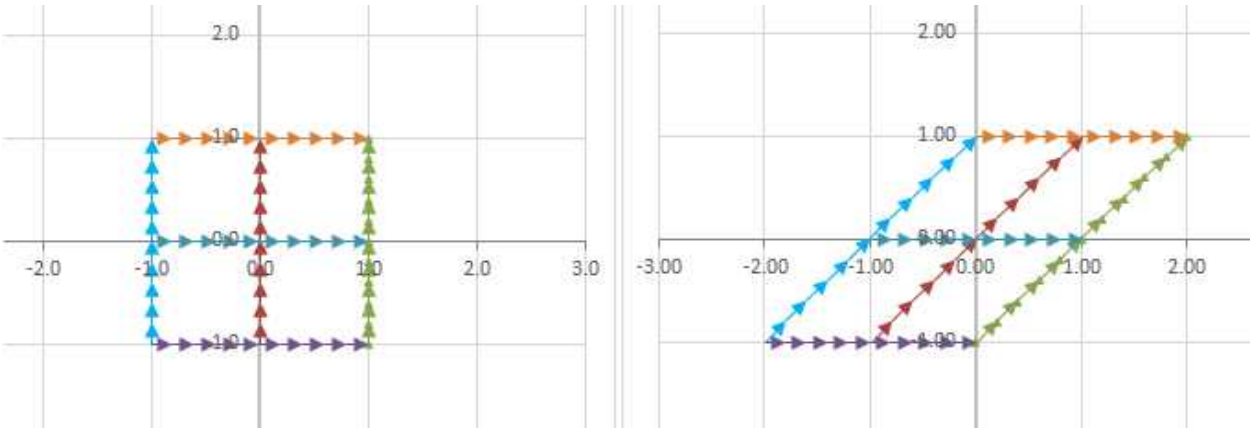
$$F = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Below, we replace circles with *ellipses* and see what happens to them under such a function:



There seems to be no stretch along the x -axis. There is still angular stretch-shrink but this time it is between the two ends of the same line.

To see more clearly, consider the Cartesian grid. This is what happens to a square:



The plane is skewed, like a deck of cards:



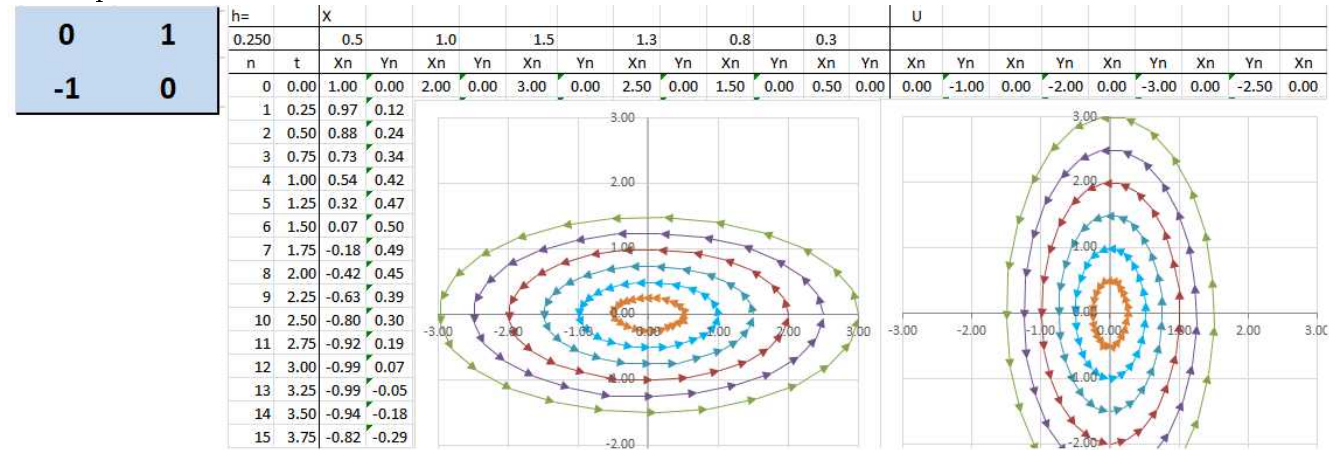
Such a skewing can be carried out with any image-editing software. The function is both one-to-one and onto.

Example 2.4.13: rotation $\pi/2$

Consider a rotation through 90 degrees: (x, y) becomes $(-y, x)$. We have:

$$\begin{cases} u = -y \\ v = x \end{cases}, \text{ re-written: } \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix}.$$

The experiment confirms what we know:



We’ve had many examples, but how do we *build* a linear operator from a description?

The solution relies on the following simple observation:

Theorem 2.4.14: Columns are Values of Basis Vectors

The two columns of the matrix of a linear operator are the values of the two basis vectors under this operator:

$$F = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \implies F \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} a \\ c \end{bmatrix} \quad \text{and} \quad F \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} b \\ d \end{bmatrix}.$$

Exercise 2.4.15

Prove the theorem.

The converse is just as important:

Theorem 2.4.16: Values of Basis Vectors Are Columns

The matrix of a linear operator is fully determined by the values of the two basis vectors under this operator.

In other words, we *merge* the two column-vectors into a matrix:

$$F : \begin{bmatrix} 1 \\ 0 \end{bmatrix} \mapsto \begin{bmatrix} a \\ c \end{bmatrix}, \quad F : \begin{bmatrix} 0 \\ 1 \end{bmatrix} \mapsto \begin{bmatrix} b \\ d \end{bmatrix} \quad \text{Merge: } F = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Example 2.4.17: matrices from values

This is the zero operator:

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} \mapsto \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 1 \end{bmatrix} \mapsto \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{Merge: } 0 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

This is the identity:

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} \mapsto \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 1 \end{bmatrix} \mapsto \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \text{Merge: } I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

This is the horizontal stretch by 2:

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} \mapsto \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 1 \end{bmatrix} \mapsto \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \text{Merge: } S_x = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}.$$

And this is the vertical stretch by 3:

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} \mapsto \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 1 \end{bmatrix} \mapsto \begin{bmatrix} 0 \\ 3 \end{bmatrix} \quad \text{Merge: } S_y = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}.$$

This is the horizontal flip:

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} \mapsto \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 1 \end{bmatrix} \mapsto \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \text{Merge: } F_x = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}.$$

And this is the vertical flip:

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} \mapsto \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 1 \end{bmatrix} \mapsto \begin{bmatrix} 0 \\ -1 \end{bmatrix} \quad \text{Merge: } F_y = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

This is the flip about the diagonal:

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} \mapsto \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 1 \end{bmatrix} \mapsto \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{Merge: } F_d = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Warning!

Its matrix is just an abbreviated representation of a linear operator.

Exercise 2.4.18

Suppose a linear operator A :

- leaves the x -axis intact, and
- stretches the y -axis by a factor of 2.

Find the matrix of A .

Exercise 2.4.19

Suppose a linear operator A :

- rotates the x -axis 45 degrees clockwise, and
- flips the y -axis.

Find the matrix of A .

Exercise 2.4.20

Suppose a linear operator A :

- leaves the x -axis intact, and
- stretches the diagonal $y = x$ by a factor of 2.

Find the matrix of A .

Exercise 2.4.21

Make up your own linear operator and find its matrix. Repeat.

Let’s apply this result to some transformations we have been interested in.

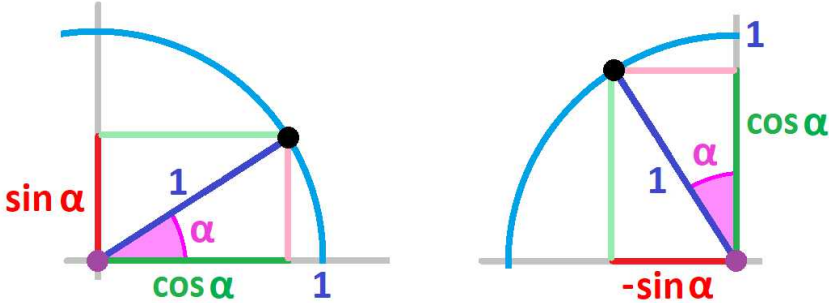
Theorem 2.4.22: Matrix of Rotation

The linear operator of rotation through an angle α is given by the following matrix:

$$R = \begin{bmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{bmatrix}$$

Proof.

We only need to see where the basis vectors $\langle 1, 0 \rangle$ and $\langle 0, 1 \rangle$ go.



The first one is simple:

$$R \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \cos \alpha \\ \sin \alpha \end{bmatrix} .$$

The second one flips the sign of the x -component:

$$R \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -\sin \alpha \\ \cos \alpha \end{bmatrix} .$$

Theorem 2.4.25: One-to-one Linear Operator

A linear operator F is one-to-one if and only if the equation $F(X) = 0$ has only the zero solution.

Proof.

Suppose there are two distinct solutions $X \neq Y$. Then, we conclude:

$$F(X) = F(Y) \implies F(X) - F(Y) = 0 \implies F(X - Y) = 0.$$

In other words, we have found such a $Z = X - Y \neq 0$ that $F(Z) = 0 = F(0)$.

Exercise 2.4.26

Prove the rest of the theorem.

In other words, F is one-to-one when

$$F(X) = 0 \implies X = 0.$$

So, to determine whether a mixture problem has a single solution, we choose, in a twist, to replace it with a mixture problem that requires to produce *zeros* in all equations. Then we ask if this problem has a non-zero solution.

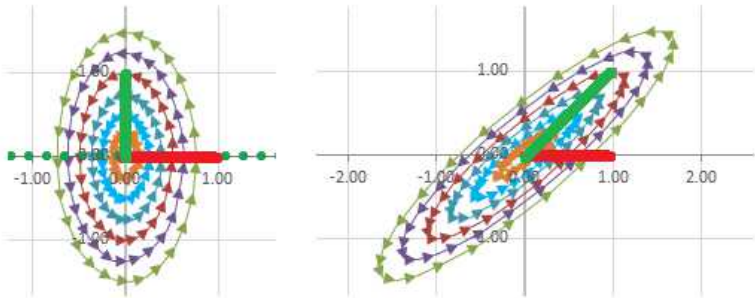
2.5. The determinant of a matrix

Example 2.5.1: one-to-one and not

Consider this matrix:

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

Is the linear operator one-to-one? Every one of our ellipses in the domain has been stretched and maybe rotated but they still cover the the whole plane in the codomain:

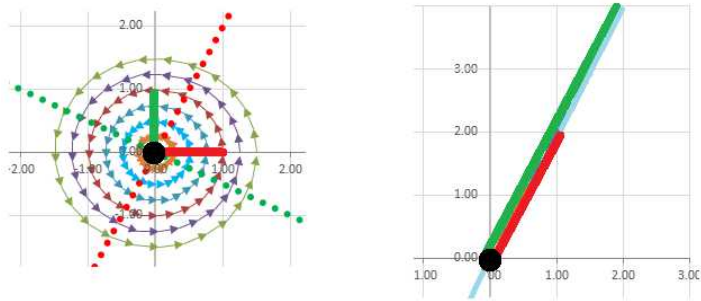


It is one-to-one.

This one is different:

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}.$$

We just watch where the two basis vectors go:



What is the difference from the former case? They go to multiples of each other: Their values are proportional to the vector $\langle 1, 2 \rangle$. We can see that in the matrix that the second column is twice the first:

$$2 \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix} .$$

In fact, a whole line of vectors goes to 0; it's not one-to-one!

Definition 2.5.2: singular matrix

A 2×2 matrix A is called *singular* when its two columns are multiples of each other.

So, for a singular matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} ,$$

there is such an x that:

$$\begin{bmatrix} a \\ c \end{bmatrix} = x \begin{bmatrix} b \\ d \end{bmatrix} .$$

Let's examine this idea: Under what circumstances is a matrix singular?

We break the vector equation above into two scalar equations:

$$a = xb, \quad c = xd .$$

Instead of solving them for x , we assume that there is such an x . We multiply the first by d , and the second by b , and then subtract to eliminate x :

$$\begin{array}{rcl} ad & = & xbd \\ cb & = & xbd \\ \hline ad - bc & = & 0 \end{array}$$

We conclude that if such an x exists, then

$$ad - bc = 0 .$$

In this expression, the terms of the matrix are cross-multiplied and subtracted:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \rightarrow ad - bc .$$

This number is an important characteristic of the matrix:

Definition 2.5.3: determinant

The *determinant* of a 2×2 matrix A is defined and denoted as follows:

$$\det A = \det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$$

What does the determinant determine?

Theorem 2.5.4: Singular Matrix and Determinant

A 2×2 matrix A is singular if and only if $\det A = 0$.

Proof.

(\Rightarrow) Suppose A is singular, then

$$\begin{bmatrix} a \\ c \end{bmatrix} = x \begin{bmatrix} b \\ d \end{bmatrix} \implies \begin{cases} a = xb \\ c = xd \end{cases} \implies \det A = ad - bc = (xb)d - b(xd) = 0.$$

(\Leftarrow) Suppose $ad - bc = 0$, then let's find x , the multiple.

- Case 1: Assume $b \neq 0$, then choose $x = \frac{a}{b}$. Then

$$\begin{aligned}xb &= \frac{a}{b}b = a, \\xd &= \frac{a}{b}d = \frac{ad}{b} = \frac{bc}{b} = c.\end{aligned}$$

So

$$x \begin{bmatrix} b \\ d \end{bmatrix} = \begin{bmatrix} a \\ c \end{bmatrix}.$$

- Case 2: Assume $a \neq 0$...

Exercise 2.5.5

Finish the proof.

We make the following observation about the determinant, which will also reappear in the case of $n \times n$ matrices:

- The determinant is an alternating sum of terms, each of which is the product of n of the matrix's entries, exactly one from each row and exactly one from each column.

Let's consider a special matrix equation, with the zero right-hand side:

$$F(X) = 0.$$

It is called a *homogeneous* equation.

We know that there is always at least one solution, the zero vector!

The question is then becomes

- Are there any non-zero solutions?

We know that the answer may be provided by a more basic question:

- Is the function one-to-one?

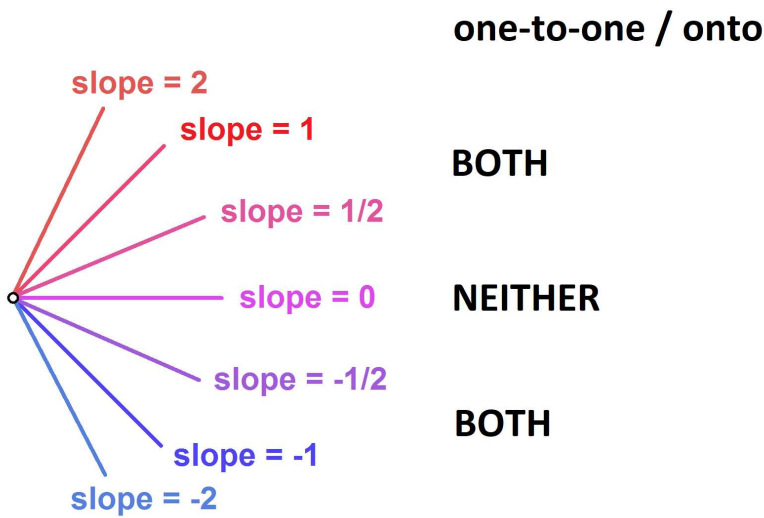
Let's start with dimension 1:

$$f(x) = mx, \text{ solve } f(x) = 0.$$

This is simple:

$$mx = 0 \implies x = 0 \text{ ... unless } m = 0.$$

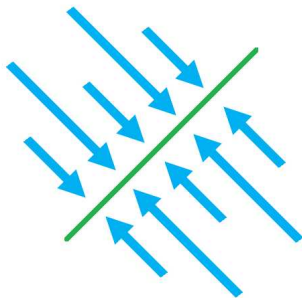
All functions except the constant zero function are one-to-one and, therefore, can only produce a single solution for our equation:



Is there a similar decisive condition in the two-dimensional case? We do have the zero operator (matrix):

$$F(X) = 0 \text{ for all } X.$$

It collapses the whole plane to the origin. However, there are other, less extreme collapses, projections onto lines:



Then a whole line is taken to 0. They are also not one-to-one!

We deploy simple algebra in order to resolve this issue.

Theorem 2.5.6: Non-zero Solutions

Suppose A is a 2×2 matrix. Then, $\det A \neq 0$ if and only if the solution set of the matrix equation $AX = 0$ consists of only 0.

Proof.

We will use the Zero Factor Property: The product of two numbers is zero if and only if either one of them (or both) is zero; i.e.,
$$a = 0 \text{ OR } b = 0 \iff ab = 0.$$

Let's solve the system of linear equations:
$$\begin{cases} ax + by = 0, & (1) \\ cx + dy = 0. & (2) \end{cases}$$

From (1), we derive:
$$y = -ax/b, \text{ provided } b \neq 0. \quad (3)$$

Substitute this into (2):
$$cx + d(-ax/b) = 0.$$

Then
$$x(c - da/b) = 0,$$

or, alternatively,

$$x(cb - da) = 0, \text{ when } b \neq 0.$$

One possibility is $x = 0$; it follows from (3) that $y = 0$ too. Then, we have two cases for $b \neq 0$:

- Case 1: $x = 0, y = 0$, or
- Case 2: $ad - bc = 0$.

Case 1 doesn't interest us. In case 2, x is arbitrary and there may be non-zero solutions.

Now, we apply this analysis to y in (1) instead of x ; we have for $a \neq 0$:

- Case 1: $x = 0, y = 0$, or
- Case 2: $ad - bc = 0$.

The result is the same! Furthermore, if we apply this analysis for x and y in (2) instead of (1), we have the same two cases. Thus, whenever one of the four coefficients, a, b, c, d , is non-zero, we have these cases:

- Case 1: $x = 0, y = 0$, or
- Case 2: $ad - bc = 0$.

But when $a = b = c = d = 0$, Case 2 is satisfied... and we can have any values for x and y !

According to the analysis above:

$$\det A \neq 0 \implies x = y = 0.$$

The converse is also true. Indeed, let's consider our system of linear equations again:

$$\begin{cases} ax + by = 0, & (1) \\ cx + dy = 0. & (2) \end{cases}$$

We multiply (1) by c and (2) by a . Then we have:

$$\begin{cases} cax + cby = 0 & c \cdot (1) \\ acx + ady = 0 & a \cdot (2) \\ \hline (ca - ac)x + (cb - ad)y = 0 \\ 0 \cdot x - \det A \cdot y = 0 \end{cases}$$

The third equation is the result of subtraction of the first two. The whole equation is zero when $\det A = 0$! This means that equations (1) and (2) represent two identical lines on the plane. It follows that the original system has infinitely many solutions.

Exercise 2.5.7

What if $a = 0$ or $c = 0$?

Example 2.5.8: computing determinants

The flip over the y -axis:

$$\det \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} = (-1) \cdot 1 - 0 \cdot 0 = -1.$$

The stretch:

$$\det \begin{bmatrix} \lambda & 0 \\ 0 & \mu \end{bmatrix} = \lambda \cdot \mu - 0 \cdot 0 = \lambda \cdot \mu.$$

The rotation:

$$\det \begin{bmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{bmatrix} = \cos \alpha \cdot \cos \alpha - (-\sin \alpha) \cdot \sin \alpha = \cos^2 \alpha + \sin^2 \alpha = 1,$$

by the Pythagorean Theorem.

These have non-zero determinants. Meanwhile, the projection on the x -axis has a zero determinant:

$$\det \begin{bmatrix} \lambda & 0 \\ 0 & 0 \end{bmatrix} = \lambda \cdot 0 - 0 \cdot 0 = 0.$$

As you can see, we can derive some more information from the value of the determinant than just that it's one-to-one.

The following is the *contra-positive* form of the theorem:

Corollary 2.5.9: Non-zero Solutions

Suppose A is a 2×2 matrix. Then, there is such an $X \neq 0$ that $AX = 0$ if and only if $\det A = 0$.

Since $A(0) = 0$, this indicates that A isn't one-to-one. There is more:

Corollary 2.5.10: Bijections and Determinants

Suppose A is a 2×2 matrix. It is a bijection if and only if $\det A \neq 0$.

Exercise 2.5.11

Prove the rest of the theorem.

So, a zero determinant indicates that some non-zero vector X is taken to 0 by A . It follows that all the multiples, kX , of X are also taken to 0:

$$A(kX) = kA(X) = k0 = 0.$$

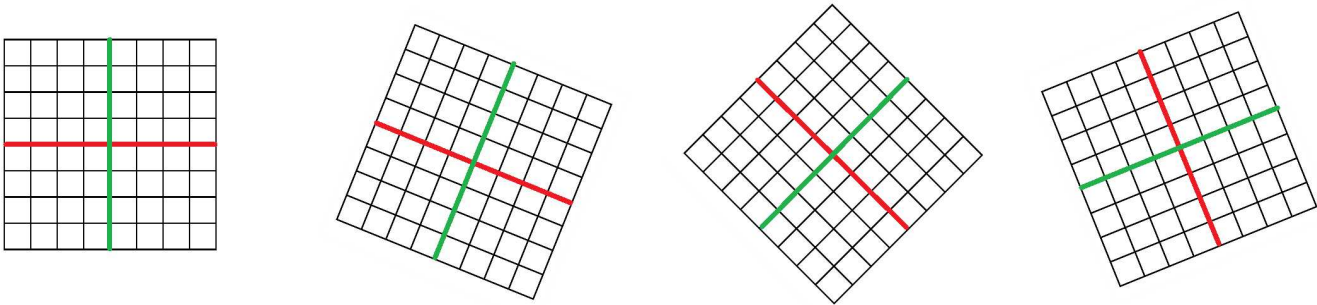
In other words, the whole line is collapsed to 0.

Theorem 2.5.12: Line Collapses

If a vector is taken to zero by a linear operator, then the whole line in the direction of this vector from the origin is taken to zero:

$$A(X) = 0 \implies A\left(\{Y : Y = kX, k \text{ real}\}\right) = 0.$$

We can place different coordinate systems on the same plane. The origin is the same, but the units and the angles of the axes may be different:



We know that the vector algebra remains the same. However, a component representation of a vector does depend on our choice of the Cartesian system. Therefore, a matrix representation of a linear operator depends on our choice of the Cartesian system too. Remarkably, this isn't true for the determinant! The following important fact is accepted without proof:

Corollary 2.5.13: Determinant Is Intrinsic

The determinant of a linear operator remains the same in any Cartesian coordinate system.

The determinant will tell us a lot about the linear operator:

- $\det A < 0$ indicates the presence of a flip.
- $|\det A| = 1$ indicates that this is a motion.
- $\det A = 0$ indicates the collapse or the presence of a projection.

But how do we detect stretches or rotations?

2.6. It’s a stretch: eigenvalues and eigenvectors

An easy observation about rotation is that if just one vector isn’t rotated, there is no rotation!

If a vector isn’t rotated, what can possibly happen to it under a linear operator? A stretch (with a possible flip). In other words, it’s scalar multiplication:

$$V \mapsto \lambda V,$$

for some real λ .

Example 2.6.1: re-scaling

Consider a linear operator given by the matrix:

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}.$$

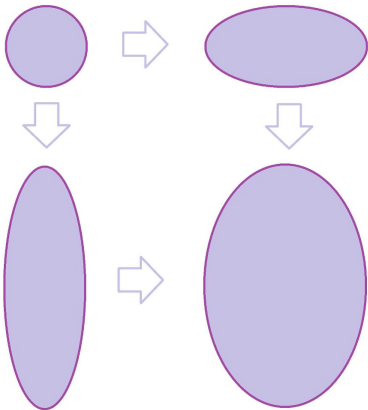
What exactly does this transformation of the plane do to it? To answer, just consider where A takes the standard basis vectors:

$$\begin{aligned} A : e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} &\mapsto \begin{bmatrix} 2 \\ 0 \end{bmatrix} = 2e_1 \\ A : e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} &\mapsto \begin{bmatrix} 0 \\ 3 \end{bmatrix} = 3e_2 \end{aligned}$$

In other words, what happens to either is a (different) scalar multiplication:

$$A(e_1) = 2e_1 \text{ and } A(e_2) = 3e_2$$

Furthermore, the entirety of each of the axes is stretched this way. So, we can say that A stretches the plane horizontally by a factor of 2 and vertically by 3, in either order:



Even though we speak of stretching the plane, this is not to say that all *vectors* are stretched. Indeed, other vectors may be rotated; for example, the values of $\langle 1, 1 \rangle$ isn't its multiple:

$$A \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \neq \lambda \begin{bmatrix} 1 \\ 1 \end{bmatrix} ,$$

for any real λ .

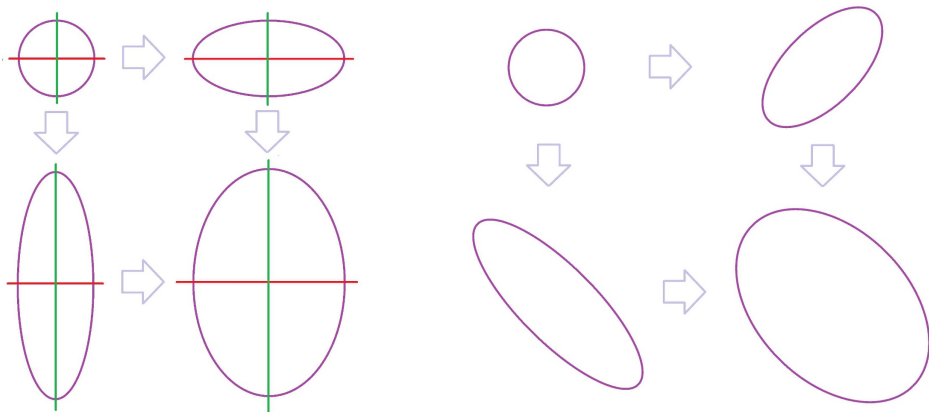
Exercise 2.6.2

Analyze a linear operator with a diagonal matrix:

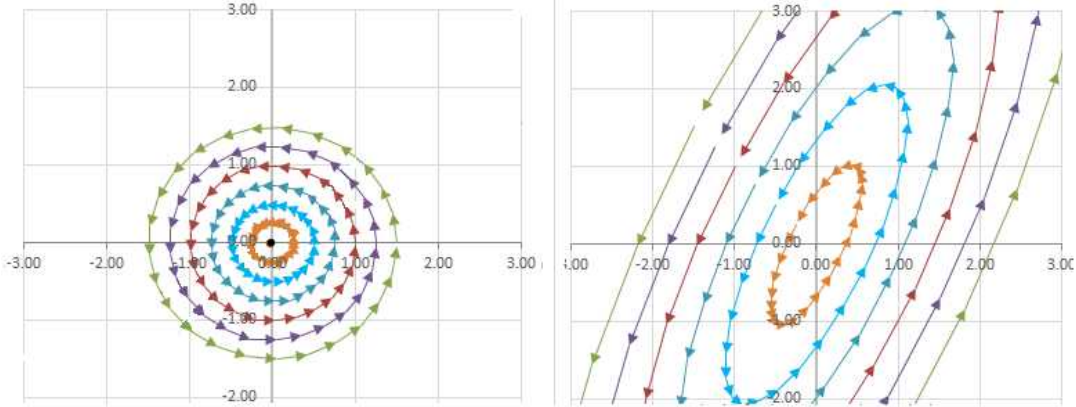
$$A = \begin{bmatrix} h & 0 \\ 0 & v \end{bmatrix} .$$

Example 2.6.3: stretch along other axes

What if the operator stretches along *other* lines? Here we simply rotate the picture in the last example through 45 degrees to make this point:



The circle is stretched, but in what direction or directions? Is there a rotation too? It is hard to tell without prior knowledge. We also have seen this:



The plane is visibly stretched, but in what direction or directions? It is hard to tell because the result simply looks skewed.

It might be typical then that a linear operator A rotates some vectors, but A also stretches other vectors. On such a vector V , A acts as a scalar multiplication:

$$A(V) = \lambda V ,$$

for some number λ . For example, we see *disproportional* horizontal and vertical stretching:



This idea bring us to the following important concept:

Definition 2.6.4: eigenvalue

Given a linear operator $A : \mathbf{R}^2 \rightarrow \mathbf{R}^2$, a (real) number λ is called an *eigenvalue* of A if it satisfies:

$$A(V) = \lambda V$$

for some non-zero vector V in \mathbf{R}^2 . Then, V is called an *eigenvector* of A corresponding to λ .

Warning!

Vector $V = 0$ is excluded because we always have $A(0) = 0$.

Note that “eigen” means “characteristic” in German.

Now, how do we find these?

Example 2.6.5: identity operator

If this is the identity matrix, $A = I$, the equation is easy to solve:

$$\lambda V = AV = IV = V.$$

So, $\lambda = 1$. This is the only eigenvalue. What are its eigenvectors? All vectors but 0. Indeed, no vector is rotated!

Exercise 2.6.6

What about a stretch by a factor of k ?

Example 2.6.7: diagonal matrix

Let's revisit this diagonal matrix:

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}.$$

Then our vector equation $AV = \lambda V$ becomes:

$$\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \lambda \begin{bmatrix} x \\ y \end{bmatrix}.$$

Let's rewrite:

$$\begin{bmatrix} 2x \\ 3y \end{bmatrix} = \begin{bmatrix} \lambda x \\ \lambda y \end{bmatrix} \implies \begin{cases} 2x = \lambda x & \text{AND} \\ 3y = \lambda y \end{cases} \implies \begin{cases} x(2 - \lambda) = 0, & (1) \\ y(3 - \lambda) = 0. & (2) \end{cases}$$

The two equations must be satisfied simultaneously.

We will use the *Zero Factor Rule* again. Now, we have $V = \langle x, y \rangle \neq 0$, so either $x \neq 0$ or $y \neq 0$. Let's use the above equations to consider these two cases:

- Case 1: $x \neq 0$, then from (1), we have: $2 - \lambda = 0 \implies \lambda = 2$.
- Case 2: $y \neq 0$, then from (2), we have: $3 - \lambda = 0 \implies \lambda = 3$.

These are the only two possibilities. We have found the eigenvalues!

The second part is to find the eigenvectors. If $\lambda = 2$, then $y = 0$ from (2). Therefore, the corresponding eigenvectors are:

$$\begin{bmatrix} x \\ 0 \end{bmatrix}, \quad x \neq 0, \quad A \begin{bmatrix} x \\ 0 \end{bmatrix} = 2 \begin{bmatrix} x \\ 0 \end{bmatrix}.$$

If $\lambda = 3$, then $x = 0$ from (1). Therefore, the corresponding eigenvectors are:

$$\begin{bmatrix} 0 \\ y \end{bmatrix}, \quad y \neq 0, \quad A \begin{bmatrix} 0 \\ y \end{bmatrix} = 3 \begin{bmatrix} 0 \\ y \end{bmatrix}.$$

These two sets are *almost* equal to the two axes! If we append 0 to these sets of eigenvectors, we have the following. For $\lambda = 2$, the set is the x -axis:

$$\left\{ \begin{bmatrix} x \\ 0 \end{bmatrix} : x \text{ real} \right\}$$

And for $\lambda = 3$, the set is the y -axis:

$$\left\{ \begin{bmatrix} 0 \\ y \end{bmatrix} : y \text{ real} \right\}$$

The solution is time-consuming, but there will be a short-cut later. We have confirmed the fact that, because of the non-uniform re-scaling, all vectors are rotated except for the vertical and horizontal ones.

Exercise 2.6.8

Analyze a linear operator with a diagonal matrix:

$$A = \begin{bmatrix} h & 0 \\ 0 & v \end{bmatrix}.$$

From the example, we can guess a pattern.

Theorem 2.6.9: Multiples of Eigenvectors

Any non-zero multiple of an eigenvector is also an eigenvector – with respect to the same eigenvalue.

Proof.

Suppose V is an eigenvector of a linear operator A corresponding to the eigenvalue λ :

$$AV = \lambda V.$$

If $W = kV$, then

AW

$= A(kV)$

$= kAV$

$= k\lambda V$

$= \lambda(kV)$

$= \lambda W$

Substitute.

Use the fact that it preserves scalar multiplication.

Use the fact that this is an eigenvector of λ .

Rearrange.

Substitute back.

The whole line is made up of eigenvectors. It's a copy of \mathbf{R} ! More general is the following:

Definition 2.6.10: eigenspace

For an eigenvalue λ of a linear operator A , the *eigenspace* of A corresponding to λ is defined and denoted by the following:

$$E(\lambda) = \{V : A(V) = \lambda V\}.$$

It's all eigenvectors of λ plus 0. We include it in order to make this set into a *space*, a vector space. From the examples above, we derive the following.

Example 2.6.11: identity matrix

For the identity matrix, we have:

$$E(1) = \mathbf{R}^2.$$

Example 2.6.12: diagonal matrix

For

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix},$$

we have:

- $E(2)$ is the x -axis.
- $E(3)$ is the y -axis.

Two copies of \mathbf{R} !

Example 2.6.13: rotation

A rotation doesn't stretch any vectors. Therefore, there are no (real) eigenvalues. Therefore, there are no eigenvectors and no eigenspaces.

Example 2.6.14: zero matrix

For the zero matrix, $A = 0$, we have:

$$AV = \lambda V, \text{ or } 0 = \lambda V.$$

Therefore, $\lambda = 0$ since $V \neq 0$. Furthermore:

$$E(0) = \mathbf{R}^2.$$

Example 2.6.15: projection

Consider now the projection on the x -axis,

$$P = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Then our matrix equation is solved as follows:

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \lambda \begin{bmatrix} x \\ y \end{bmatrix} \implies \begin{bmatrix} x \\ 0 \end{bmatrix} = \begin{bmatrix} \lambda x \\ \lambda y \end{bmatrix} \implies \begin{cases} x = \lambda x \text{ AND} \\ 0 = \lambda y \end{cases}$$

So, the only possible cases are:

$$\lambda = 0 \text{ and } \lambda = 1.$$

It appears that the operator is projecting in one direction and doing nothing in another.

Next, in order to find the corresponding eigenvectors, we now go back to the system of linear equations for x and y . We consider these two cases. First:

Case 1: $\lambda = 0 \implies \begin{cases} x = 0 \cdot x \text{ AND} \\ 0 = 0 \cdot y \end{cases} \implies \begin{cases} x = 0 \text{ AND} \\ y \text{ any} \end{cases} \implies E(0) = \left\{ \begin{bmatrix} 0 \\ y \end{bmatrix} : y \text{ real} \right\}$

This is the y -axis. Second:

Case 2: $\lambda = 1 \implies \begin{cases} x = 1 \cdot x \text{ AND} \\ 0 = 1 \cdot y \end{cases} \implies \begin{cases} x \text{ any AND} \\ y = 0 \end{cases} \implies E(1) = \left\{ \begin{bmatrix} x \\ 0 \end{bmatrix} : x \text{ real} \right\}$

This is the x -axis.

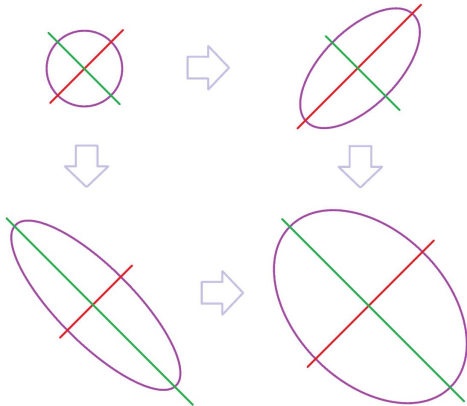
Typically, we have two eigenvectors that aren't multiples of each other:

- $A(V_1) = \lambda_1 V_1$ and
- $A(V_2) = \lambda_2 V_2$,

for some numbers $\lambda_1 \neq \lambda_2$.

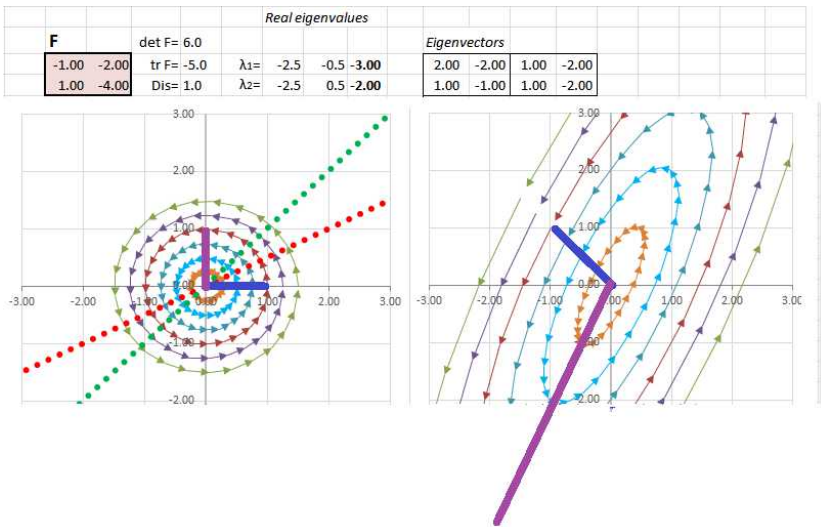
Example 2.6.16: stretch along other axes

Let's revisit the stretch along special lines. It might look like this:

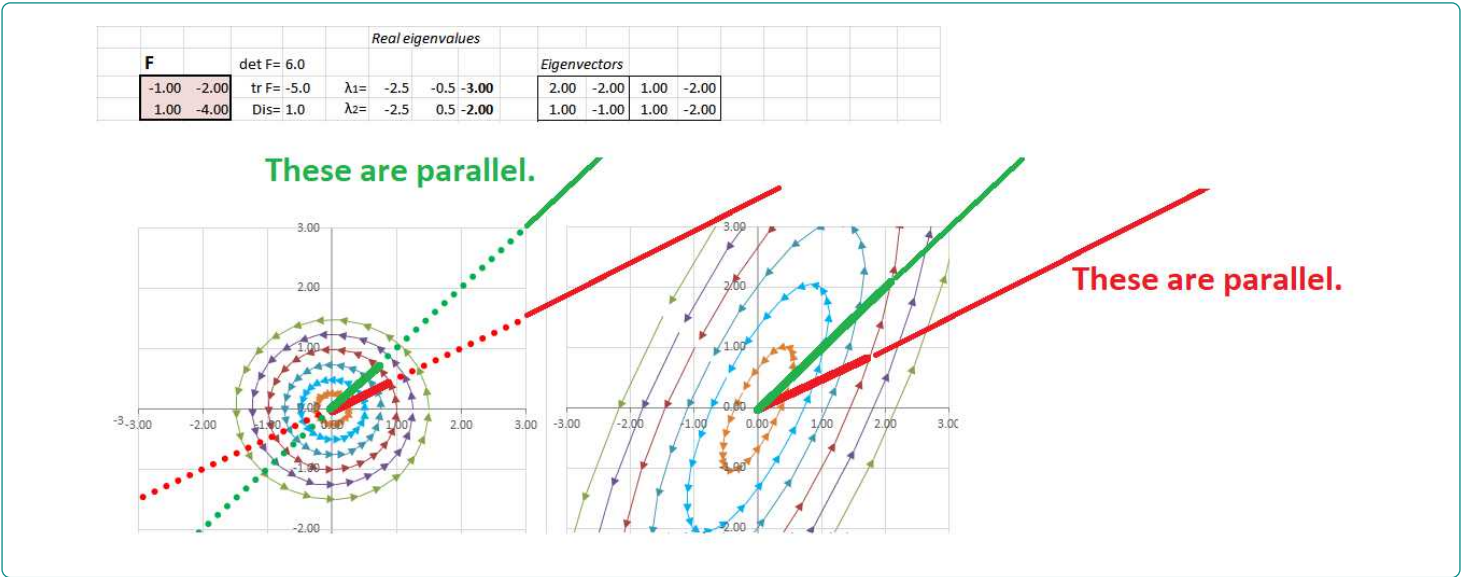


However, how would we even find these special directions?

Below, we try the basis vectors, look at where they go (from the matrix itself), and see that they have rotated:



They are not eigenvectors! The eigenvectors below may be found by trial and error or by the method presented below; they, indeed, don't rotate:



Example 2.6.17: no eigenvectors

Can we derive what a linear operator does from its matrix only, without visualization? Suppose A is given:

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} .$$

Then, to find the eigenvalues, we consider this system of linear equations:

$$AV = \lambda V \implies \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \lambda \begin{bmatrix} x \\ y \end{bmatrix} .$$

We solve it as follows:

$$\implies \begin{cases} -y = \lambda x \\ x = \lambda y \end{cases} \text{ AND } \implies \begin{cases} -xy = \lambda x^2 \\ xy = \lambda y^2 \end{cases} \text{ AND } \implies \lambda x^2 = -\lambda y^2 \implies x^2 = -y^2 \text{ OR } \lambda = 0 .$$

A direct examination reveals that $\lambda = 0$ is *not* an eigenvalue:

$$AV = 0 \cdot V \implies \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0 \implies x = 0, y = 0 .$$

In the meantime, the equation $x^2 = -y^2$ is impossible unless both x and y are zeros, which is not allowed.

There seems to be no eigenvalues, certainly not *real* ones... This means that every vector is rotated. Maybe this *is* a rotation? Yes, we recognize the matrix of the 90-degree rotation.

Exercise 2.6.18

How does the determinant of A tell you whether 0 is an eigenvalue?

Exercise 2.6.19

Show that a zero eigenvalue implies a collapse.

Example 2.6.20: need for homogeneous system

Let's revisit this linear operator:

$$A = \begin{bmatrix} -1 & -2 \\ 1 & -4 \end{bmatrix} .$$

Our vector equation becomes:

$$\begin{bmatrix} -1 & -2 \\ 1 & -4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \lambda \begin{bmatrix} x \\ y \end{bmatrix} .$$

We rewrite, again, as a system of linear equations:

$$\begin{bmatrix} -x - 2y \\ x - 4y \end{bmatrix} = \begin{bmatrix} \lambda x \\ \lambda y \end{bmatrix} \implies \begin{cases} (-1 - \lambda)x - 2y = 0 & \text{AND} \\ x + (-4 - \lambda)y = 0 \end{cases}$$

This is another system of linear equations to be solved, again. It is more complex than the ones we saw above and none of the shortcuts are available... The system corresponds to a *homogeneous* vector equation:

$$\begin{bmatrix} -1 - \lambda & -2 \\ 1 & -4 - \lambda \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} .$$

What do we know about those?

Let's review.

Suppose we have a linear operator A and we need to find its eigenvalues and eigenvectors. Let's, for now, concentrate on the former. Suppose λ is an eigenvalue of A . This means that λ is a real number and there is some non-zero vector V that satisfies:

$AV = \lambda V$

Let's do some vector algebra:

$$AV = \lambda V \implies AV - \lambda V = 0 .$$

We want to turn this equation of vectors and matrices into one entirely of matrices. We can take this equation one step further by observing that

$$\lambda V = \lambda IV ,$$

where I is the identity matrix. The linearity of these operators allows to factor V out of our equation. It takes a new form:

$(A - \lambda I)V = 0$

The equation characterizes an eigenvector and its eigenvalues in a space of any dimension.

Example 2.6.21: dimension 2

In the \mathbf{R}^2 case, we make this specific when our linear operator A is specific. Suppose it is given by a matrix:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} .$$

We carry out these computations:

$$AV = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix} \quad \text{and} \quad \lambda \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \lambda x \\ \lambda y \end{bmatrix} .$$

These two vectors are supposed to be equal, so we have a system of linear equations, which is then transformed into a *homogeneous* form:

$$\begin{cases} ax + by = \lambda x & \text{AND} \\ cx + dy = \lambda y. \end{cases} \iff \begin{cases} (a - \lambda)x + by = 0 & \text{AND} \\ cx + (d - \lambda)y = 0. \end{cases}$$

The matrix of this system is:

$$G = \begin{bmatrix} a - \lambda & b \\ c & d - \lambda \end{bmatrix} .$$

Following these computations, we recognize some matrix algebra:

$$G = \begin{bmatrix} a & b \\ c & d \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

We go back to our compact representation:

Theorem 2.6.22: Eigenvalues and Eigenvectors

Suppose A is a linear operator. Then every pair of an eigenvalue λ and its eigenvector V of A satisfy the following matrix equation:

$$GV = 0, \text{ where } G = A - \lambda I$$

Now, the question about the eigenvalues of the matrix A becomes one about the matrix G :

- Under what circumstances does the system $GV = 0$ have a non-zero solution?

We know the answer from the last section:

- The system $GV = 0$ has a non-zero solution if and only if $\det G = 0$.

We have proven the following result:

Theorem 2.6.23: Eigenvalues as Roots

Suppose A is a linear operator \mathbf{R}^2 . Then every eigenvalue λ of A is a solution to the following equation:

$$\det(A - \lambda I) = 0$$

In contrast to the *matrix* equation in the last theorem, this simple *algebraic* equation allows to discover eigenvalues and then, possibly, their eigenvectors.

We codify this idea below:

Definition 2.6.24: characteristic polynomial

The *characteristic polynomial* of a 2×2 matrix A is defined to be:

$$\chi_A(\lambda) = \det(A - \lambda I)$$

Meanwhile, the equation $\chi_A(\lambda) = 0$ is called the *characteristic equation*.

This is the convenient form of the equation we are to solve for dimension $n = 2$:

$$\chi_A(\lambda) = \det \begin{bmatrix} a - \lambda & b \\ c & d - \lambda \end{bmatrix} = (a - \lambda)(d - \lambda) - bc = 0.$$

It's *quadratic*!

We don't know what a linear operator does but – even without the eigenvectors – we can tell a lot from its eigenvalues. We rediscover some of the information about familiar operators below.

Example 2.6.25: re-scaling?

Consider again:

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} .$$

Then we solve:

$$\chi_A(\lambda) = \det \begin{bmatrix} 2 - \lambda & 0 \\ 0 & 3 - \lambda \end{bmatrix} = (2 - \lambda)(3 - \lambda) = 0 .$$

Therefore, we have:

$$\lambda = 2, \, 3 .$$

We conclude that the linear operator *stretches the plane by these factors in two different directions*. What are those directions? We can't tell without finding the eigenvectors.

Example 2.6.26: projection?

If

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} ,$$

the characteristic equation is:

$$\chi_A(\lambda) = \det \begin{bmatrix} 1 - \lambda & 0 \\ 0 & -\lambda \end{bmatrix} = (1 - \lambda)(-\lambda) = 0 .$$

Then,

$$\lambda = 1, \, 0 .$$

We conclude that the linear operator *does nothing in one direction and collapses in another*. That's a projection! What are those directions? We don't know without the eigenvectors.

Example 2.6.27: rotation?

Consider

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} .$$

Then,

$$\chi_A(\lambda) = \det \begin{bmatrix} -\lambda & -1 \\ 1 & -\lambda \end{bmatrix} = \lambda^2 + 1 = 0 .$$

No real solutions! So, no non-zero vector is taken by A to its own multiple. Maybe this is a rotation...

These three examples suggest a classification of linear operators of the plane. But first a quick review of quadratic polynomials.

Consider one:

$$f(x) = x^2 + px + q .$$

The *Quadratic Formula* then provides the x -intercepts of this function:

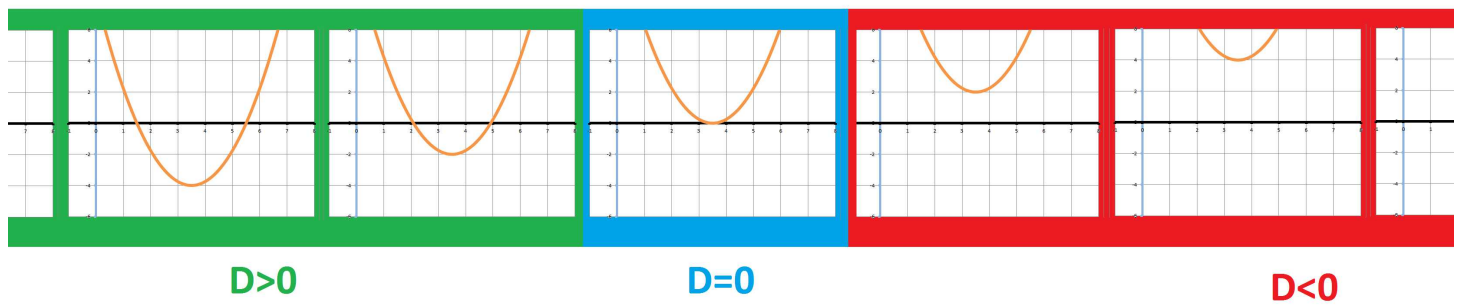
$$x = -\frac{p}{2} \pm \frac{\sqrt{p^2 - 4q}}{2} .$$

Of course, the x -intercepts are the real solutions of this equation and that is why the result only makes sense when the *discriminant* of the quadratic polynomial,

$$D = p^2 - 4q ,$$

is non-negative.

Increasing the value of the free term q makes the graph of $y = f(x)$ shift upward and, eventually, pass the x -axis entirely. We can observe how its two x -intercepts start to get closer to each other, then merge, and finally disappear:



This process is explained by what is happening, with the growth of q , to the roots given by the *Quadratic Formula*:

$$x_{1,2} = -\frac{p}{2} \pm \frac{\sqrt{D}}{2} .$$

There are three states:

1. Starting with a positive value, D decreases, and $\frac{\sqrt{D}}{2}$ decreases.
2. Then D becomes 0 and, therefore, we have $\frac{\sqrt{D}}{2} = 0$.
3. Then D becomes negative, and there are no real roots (complex roots are discussed in the next chapter).

So, we have:

- The eigenvalues are the real roots of the (quadratic) characteristic polynomial χ_A .
- Therefore, the number of eigenvalues is less than or equal to 2, counting their multiplicities.

Let's try to expand the characterstic polynomial and see if patterns emerge:

$$\begin{aligned} \chi(\lambda) &= \det \begin{bmatrix} a - \lambda & b \\ c & d - \lambda \end{bmatrix} \\ &= (a - \lambda)(d - \lambda) - bc \\ &= ad - a\lambda - \lambda d + \lambda^2 - bc \\ &= \lambda^2 - (a + d)\lambda + (ad - bc) \\ &= \lambda^2 - \operatorname{tr} A \lambda + \det A . \end{aligned}$$

The term in the middle is defined as follows:

Definition 2.6.28: trace of matrix

The *trace* of a matrix A is the sum of its diagonal elements. It is denoted by:

$$\operatorname{tr} A = \operatorname{tr} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = a + d$$

So, the trace appears – along with the determinant – in the characteristic polynomial:

Theorem 2.6.29: Characteristic Polynomial

The characteristic polynomial of matrix A takes this form:

$$\chi_A(\lambda) = \lambda^2 - \operatorname{tr} A \cdot \lambda + \det A$$

It is known that not only the determinant but also the trace are independent of our choice of a Cartesian system. Therefore, so is the characteristic polynomial.

The discriminant of the characteristic polynomial can now be used to tell what the linear operators does:

$$D = (\operatorname{tr} A)^2 - 4 \det A.$$

Theorem 2.6.30: Classification of Linear Operators

Suppose A is a linear operator given by a 2×2 matrix and D is the discriminant of its characteristic polynomial. Then we have three cases:

- 1. $D > 0$. The eigenvalues are distinct: Operator A non-uniformly re-scales the plane in the distinct directions of the corresponding eigenvectors.
- 2. $D = 0$. The eigenvalues are equal: Operator A uniformly re-scales the plane in all directions unless the eigenvectors are all multiples of each other.
- 3. $D < 0$. There are no eigenvalues: Operator A rotates the plane (with a possible re-scaling).

Proof.

For Part 1, we already know that the directions are distinct because if two eigenvectors are multiples of each other, then they have the same eigenvalue. Indeed:

$$A(V) = \lambda V \implies A(kV) = kA(V) = k\lambda V = \lambda(kV).$$

Parts 2 and 3 are addressed later in the chapter.

2.7. The significance of eigenvectors

We have shown how one can *visualize* the way a linear operator transforms the plane: by examining what happens to various curves in the domain. By mapping these curves, one can discover stretching, shrinking in various directions, rotations, etc.

In the last section, we also saw how one can *understand* the way a linear operator transforms the plane: by examining its eigenvalues. The method is entirely algebraic rather than experimental. We simply find the directions of pure stretch for F :

$$FV = \lambda V$$

The visualizations are produced by a spreadsheet. The spreadsheet also computes the eigenvector and its eigenvalues; they are shown above the graphs. The spreadsheet also shows eigenspaces as two (or one, or none) straight lines; they remain in place under the transformation.

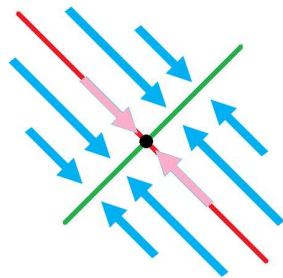
We would like to learn how to predict the outcome by examining only its matrix.

Below is a familiar fact that will take us down that road:

Theorem 2.7.1: Preimages of Zero

If the image of $V \neq 0$ under a linear operator F is zero, then so is that of any of its multiples kV .

In other words, the whole line with V as its direction vector is collapsed to 0 by F :



Example 2.7.2: collapse on axis

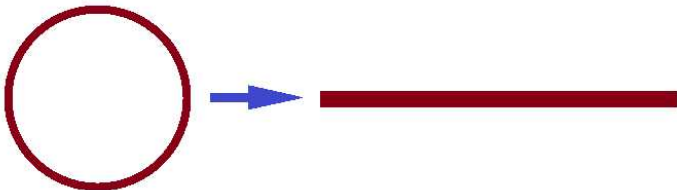
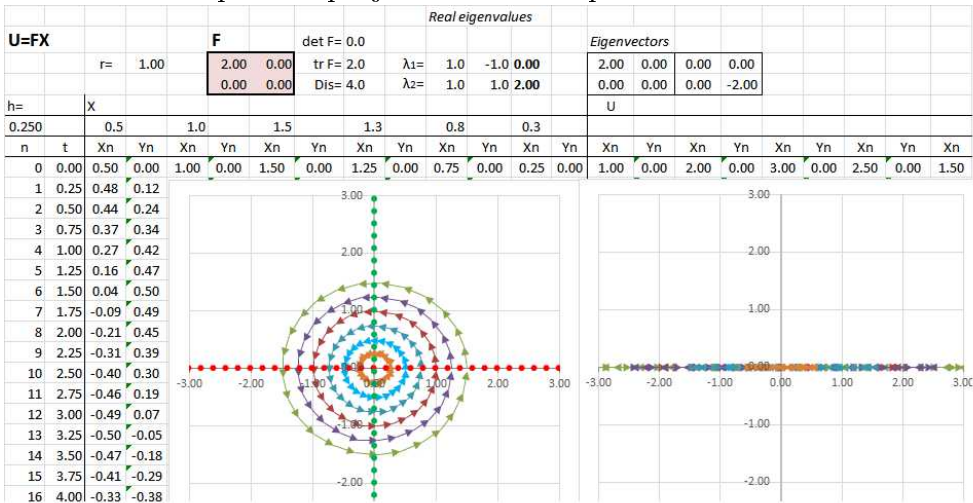
We start with a familiar example:

$$\begin{cases} u = 2x \\ v = 0 \end{cases} \text{ is re-written as } \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

Even without the characteristic equation, we can guess the eigenvalue-eigenvector pairs:

$$\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0 \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Below, one can see how this operator projects the whole plane to the x -axis:



The operator collapses the y -axis to 0, while the x -axis is stretched by a factor of 2. The standard basis vectors happen to be eigenvectors! That's the reason why the matrix is so simple.

Example 2.7.3: stretch-shrink along axes

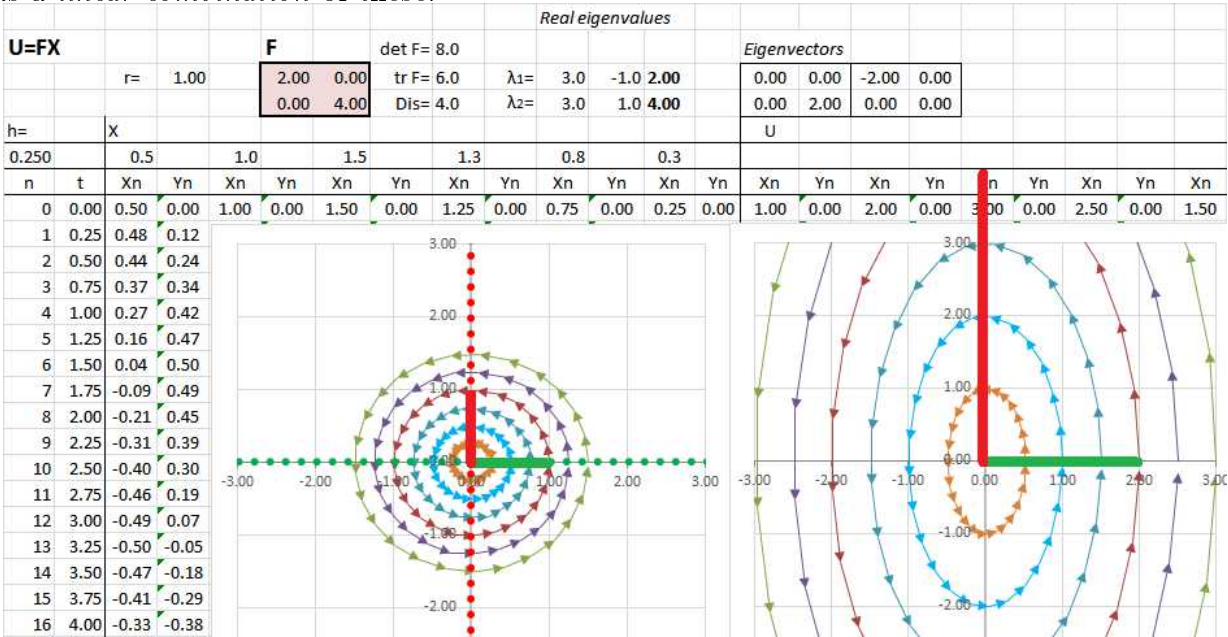
Let’s consider this linear operator and its matrix

$$\begin{cases} u &= 2x \\ v &= 4y \end{cases} \quad \text{and} \quad F = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} .$$

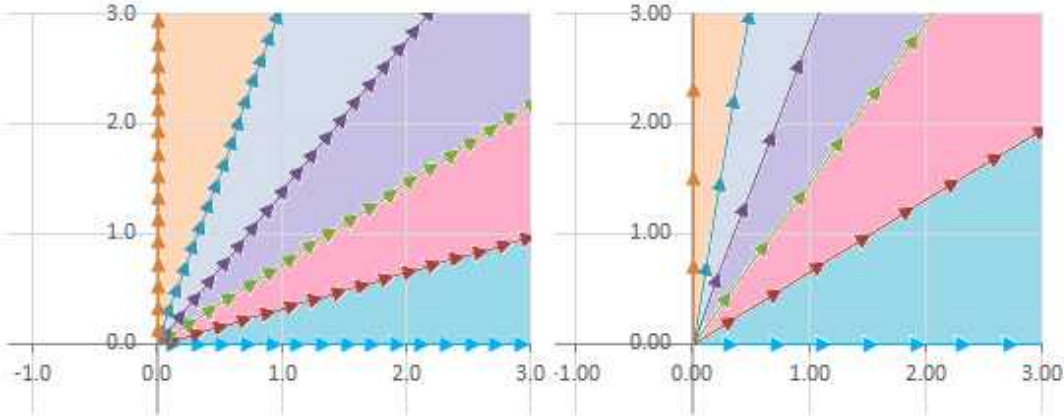
Once again, we don’t need the characteristic equation to suggest the eigenvalues (and then find the eigenvectors):

$$\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 4 \begin{bmatrix} 0 \\ 1 \end{bmatrix} .$$

As it turns out, we only need to track the values of the basis vectors, and the rest of the values are seen as a *linear combination* of these:



The rest of the vectors turn non-uniformly; i.e., they “fan out”:



This is what happens to an arbitrary vector $X = \langle x, y \rangle$:

$$FX = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = x \begin{bmatrix} 2 \\ 0 \end{bmatrix} + y \begin{bmatrix} 0 \\ 4 \end{bmatrix} = 2x \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 4y \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 2xe_1 + 4ye_2 .$$

The last expression is a linear combination of the values of the two *standard* basis vectors. The middle, however, is also a linear combination but with respect to these two vectors:

$$V_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \quad \text{and} \quad V_2 = \begin{bmatrix} 0 \\ 4 \end{bmatrix} .$$

They can be thought of as forming a “non-standard” basis. Though not unit vectors as the standard ones, they are still aligned with the axes. Now, what is the point? Every vector can be expressed as

a linear combination of the two:

$$\langle x,y \rangle = \frac{x}{2}V_1 + \frac{y}{4}V_2 .$$

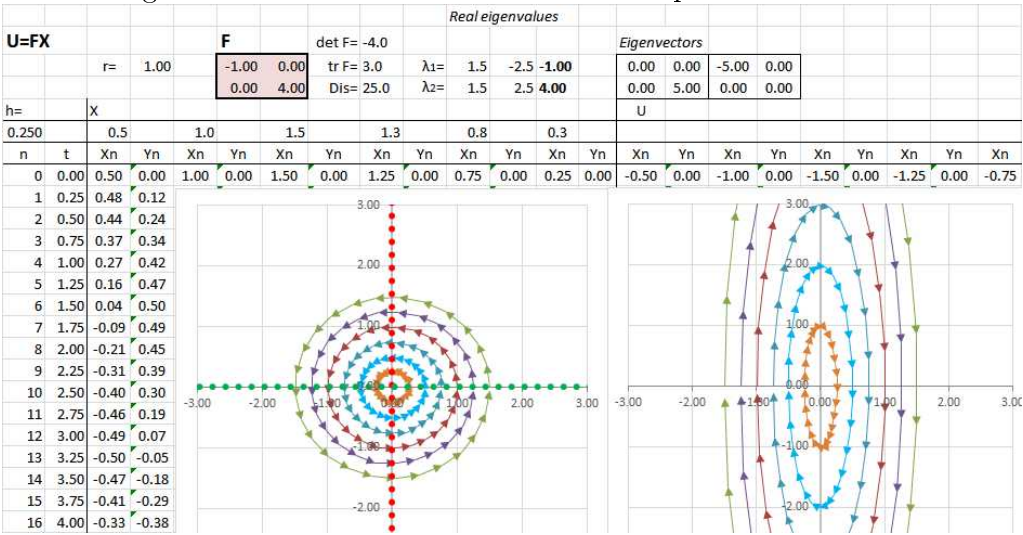
Furthermore, to know where any X goes under F , we need only to know where these two go: It’s pure scalar multiplication! We will see that any pair of eigenvectors – when not multiples of each other – would do.

Example 2.7.4: stretch-shrink along axes

A slightly different operator is the following:

$$\begin{cases} u &= -x \\ v &= 4y \end{cases} \quad \text{and} \quad F = \begin{bmatrix} -1 & 0 \\ 0 & 4 \end{bmatrix} .$$

It is still simple because the two variables remain fully separated. As a result, the two transformations of the axes can be thought of as transformations of the whole plane:



The negative sign produces a different pattern: We see the reversal of the direction of the ellipse around the origin. Algebraically, we have as before:

$$FX = \begin{bmatrix} -1 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = x \begin{bmatrix} -1 \\ 0 \end{bmatrix} + y \begin{bmatrix} 0 \\ 4 \end{bmatrix} = -x \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 4y \begin{bmatrix} 0 \\ 1 \end{bmatrix} = -xe_1 + 4ye_2 .$$

Once again, the last expression is a linear combination of the values of the two *standard* basis vectors, while the middle is a linear combination but with respect to *another basis* made up of the eigenvectors:

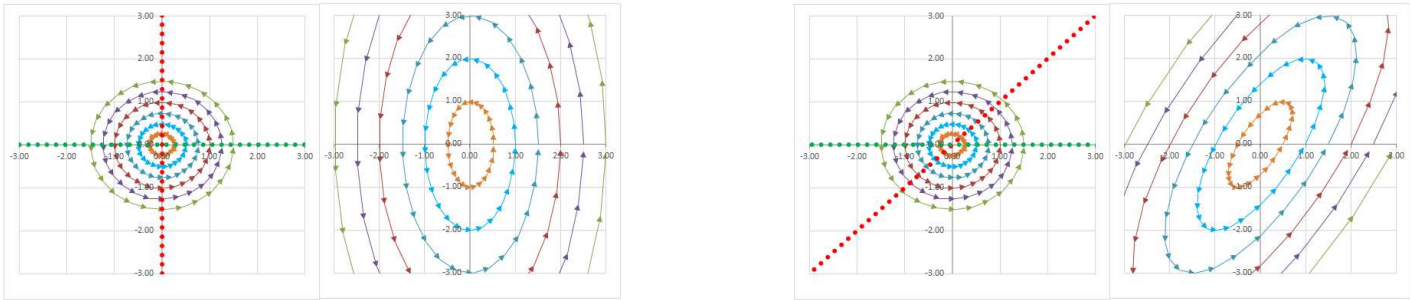
$$V_1 = \begin{bmatrix} -1 \\ 0 \end{bmatrix} \quad \text{and} \quad V_2 = \begin{bmatrix} 0 \\ 4 \end{bmatrix} .$$

What if the matrix isn’t diagonal?

Exercise 2.7.5

Show that the standard basis vectors e_1, e_2 are eigenvectors of diagonal matrices.

Instead of the standard basis vectors, we will concentrate on the eigenvectors of the operator in order to understand what the operator does. To illustrate, just imagine that the picture on the left has been skewed, resulting in the image of the right:

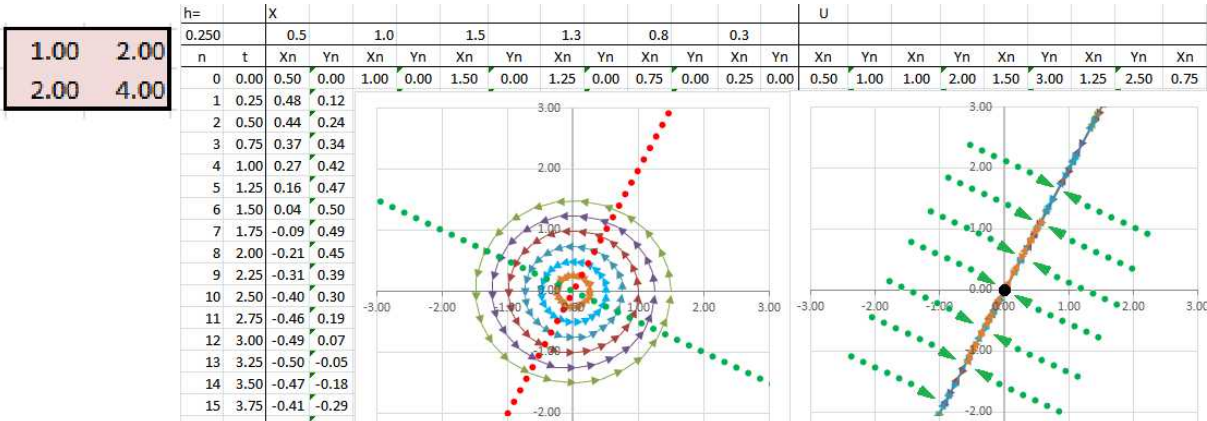


The eigenvectors of the matrix will serve as an alternative basis.

Example 2.7.6: collapse

Let’s consider a more general linear operator:

$$\begin{cases} u &= x &+& 2y \\ v &= 2x &+& 4y \end{cases} \implies F = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}.$$



It appears that the function has a stretching in one direction and a collapse in another. What are those directions? Linear algebra gives the answer.

Even without looking for eigenvectors, we know that we can use the fact that *the determinant is zero*:

$$\det F = \det \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} = 1 \cdot 4 - 2 \cdot 2 = 0.$$

It’s not one-to-one and, in fact, there is a whole line of points X with $FX = 0$. To find it, we solve this equation by solving this system of equations:

$$\begin{cases} x &+& 2y &= 0 \\ 2x &+& 4y &= 0 \end{cases} \implies x = -2y.$$

The two equations are equivalent and represent the same line. We have, indirectly, found the eigenspace and, of course, the eigenvectors of the zero eigenvalue $\lambda_1 = 0$. We can take this eigenvector for further use:

$$V_1 = \begin{bmatrix} 2 \\ -1 \end{bmatrix} \implies FV_1 = 0.$$

Let’s instead turn to the characteristic polynomial:

$$\det(F - \lambda I) = \det \begin{bmatrix} 1 - \lambda & 2 \\ 2 & 4 - \lambda \end{bmatrix} = \lambda^2 - 5\lambda = \lambda(\lambda - 5) \implies \lambda_1 = 0, \lambda_2 = 5.$$

Let’s find the eigenvectors for $\lambda_2 = 5$. We need to solve the vector equation:

$$FV = 5V,$$

i.e.,

$$FV = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 5 \begin{bmatrix} x \\ y \end{bmatrix} .$$

This gives as the following system of two linear equations (it’s the same equation):

$$\begin{cases} x + 2y = 5x \text{ AND} \\ 2x + 4y = 5y \end{cases} \implies \begin{cases} -4x + 2y = 0 \text{ AND} \\ 2x - y = 0 \end{cases} \implies y = 2x .$$

This line is the eigenspace. We choose a vector along this line to as the eigenvector for further use:

$$V_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} .$$

We summarize what F does:

- A projection along the vector $\langle 2, -1 \rangle$: The line $x = -2y$ is collapsed to 0.
- A stretch by a factor of 5 along the vector $\langle 1, 2 \rangle$: The line $y = 2x$ is stretched without any rotation.

We have confirmed the illustration above!

Furthermore, the two eigenvectors aren’t multiples of each other. That is why every vector is a linear combination of the two eigenvectors and, therefore, its value under F is a linear combination of the eigenvectors too:

$$X = xV_1 + yV_2 \implies FX = x \cdot 0 \cdot V_1 + y \cdot 5 \cdot V_2 .$$

We derive where X goes from the above summary!

Exercise 2.7.7

Find the line of the projection.

We are able to summarize what the operator does from the algebra only. The idea is uncomplicated:

- The linear operator *within the eigenspace* is “1-dimensional”; it can then be represented by a single number.

This number, the stretch-shrink factor, is of course the eigenvalue.

If we know these two numbers, how do we find the rest of the values of the linear operator? In two steps.

First, every X that lies within the eigenspace, which is a line, is a multiple of the eigenvector, and its value under F can be easily computed:

$$X = rV \implies U = F(rV) = rFV = r\lambda V .$$

Second, the rest of the values are found by following this idea: Try to express the value as a linear combination of two values found so far.

Let’s provide a foundation for this idea.

2.8. Bases

The matrix representation of a linear operator is determined by our choice of the Cartesian system. On the other hand, what it does may be described with such words as “stretch”, “rotation”, “flip”, etc. These

descriptions have nothing to do with the coordinate system. And neither do such algebraic characteristics of the operator as the trace, the determinant, the eigenvalues and the eigenvectors. We are on the right track! Once again, dealing with *vectors* instead of points requires a different approach. The *standard basis* of \mathbf{R}^2 , as before, consists of these two:

$$e_1 = \langle 1, 0 \rangle, \quad e_2 = \langle 0, 1 \rangle .$$

The components of a vector $X = \langle a, b \rangle$ with respect to this basis are, as before, a, b :

$$\langle a, b \rangle = a \langle 1, 0 \rangle + b \langle 0, 1 \rangle = ae_1 + be_2 .$$

In other words, every vector can be represented as a *linear combination* of these two vectors. However, they aren't the only ones with this property! For example, let's rewrite the above representation:

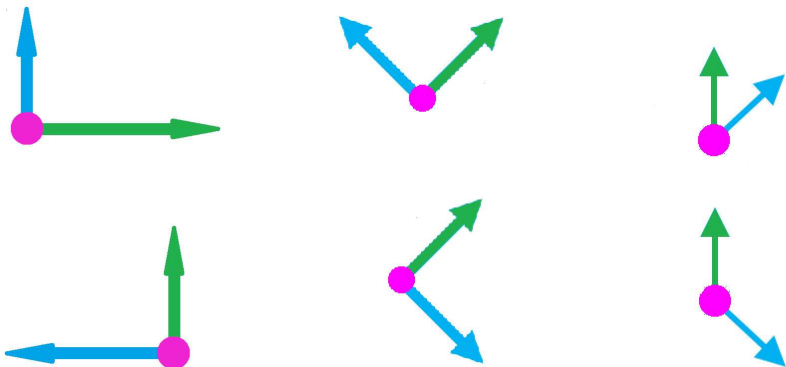
$$\langle a, b \rangle = a \langle 1, 0 \rangle + b \langle 0, 1 \rangle = ae_1 + (-b)(-e_2) .$$

So, this vector is a linear combination of the two vectors $V_1 = e_1$ and $V_2 = -e_2$.

Exercise 2.8.1

Represent vector $\langle a, b \rangle$ in terms of e_1 and $e_1 + e_2$.

All of these pairs of vectors may serve in such representations:



Exercise 2.8.2

Show that vectors that are multiples of each other can't be used for this purpose.

A very important concept below captures this idea:

Definition 2.8.3: basis

A *basis* of \mathbf{R}^2 is any such pair of vectors V_1, V_2 that every vector can be represented as a linear combination of these vectors:

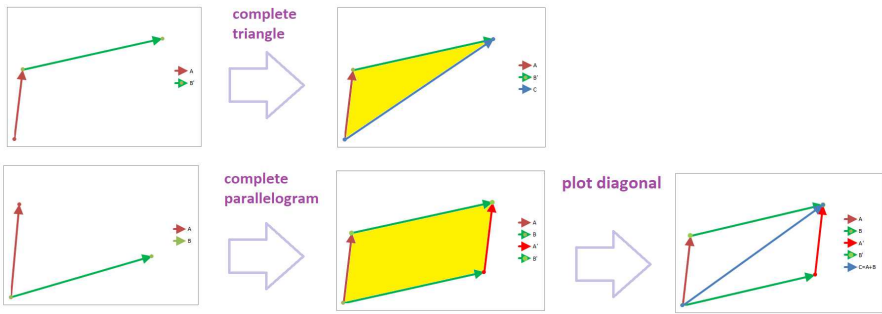
$$X = r_1V_1 + r_2V_2 .$$

Then the coefficients r_1, r_2 are called the *components* of X with respect to the basis.

Exercise 2.8.4

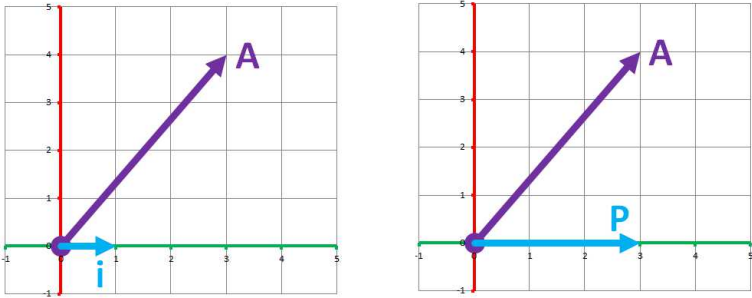
Prove that there is only one such representation.

The reasoning is that the algebra of vectors was established *before* the Cartesian system was added to the vector space and before the two operations were expressed in terms of the components of vectors. For example, this is how we add vectors:

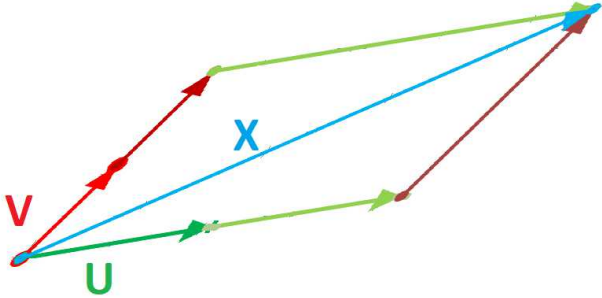


Example 2.8.5: components

The components of a vector in terms of the standard basis are found via the orthogonal projections:



A very different choice of basis U, V is shown below:



Here, vector X has components 2 and 2 with respect to this basis:

$$X = 2U + 2V .$$

Example 2.8.6: component algebra

Linear combinations are behind the components algebra. Collecting common terms after addition is what happens to the components:

	components	linear combination
U	$= \langle 2, 3 \rangle$	$= 2e_1 + 3e_2$
W	$= \langle -1, 2 \rangle$	$= -1e_1 + 2e_2$
$U + V$	$= \langle 2, 3 \rangle + \langle -1, 2 \rangle$	$= (2e_1 + 3e_2) + (-1e_1 + 2e_2)$
	$= \langle 2 - 1, 3 + 2 \rangle$	$= (2e_1 - 1e_1) + (3e_2 + 2e_2)$
	$= \langle 1, 5 \rangle$	$= 1e_1 + 5e_2$

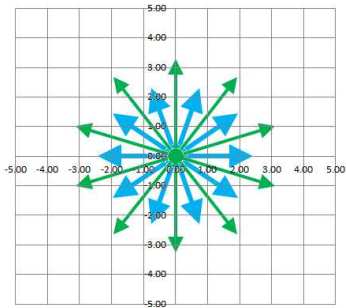
Replacing e_1, e_2 with, say, V_1, V_2 won't change anything in this computation. Same for scalar multi-

plication:

	components	linear combination
U	$= < 2, 3 >$	$= 2e_1 + 3e_2$
r	$= -3$	
rU	$= (-3) < 2, 3 >$	$= (-3)(2e_1 + 3e_2)$
	$= < (-3)2, (-3)3 >$	$= (-3)2e_1 + (-3)(1e_1)$
	$= < -6, -9 >$	$= -6e_1 - 9e_2$

Example 2.8.7: non-basis

We need to cover all the vectors:



Let’s try $V_1 = < 1, 0 >$, $V_2 = < 2, 0 >$. What are all possible linear combinations? For all pairs r_1, r_2 , we have

$$X = r_1V_1 + r_2V_2 = r_1 < 1, 0 > + r_2 < 2, 0 > = < r_1 + 2r_2, 0 > .$$

The second component will remain 0 no matter what the coefficients are! This is not a basis because we can’t represent some of the vectors.

The general result is as follows:

Theorem 2.8.8: Basis on the Plane

Any two vectors that aren’t multiples of each other and only they form a basis of \mathbf{R}^2 ; i.e.,

$$V_1 = rV_2 \iff \{V_1, V_2\} \text{ is not a basis.}$$

Exercise 2.8.9

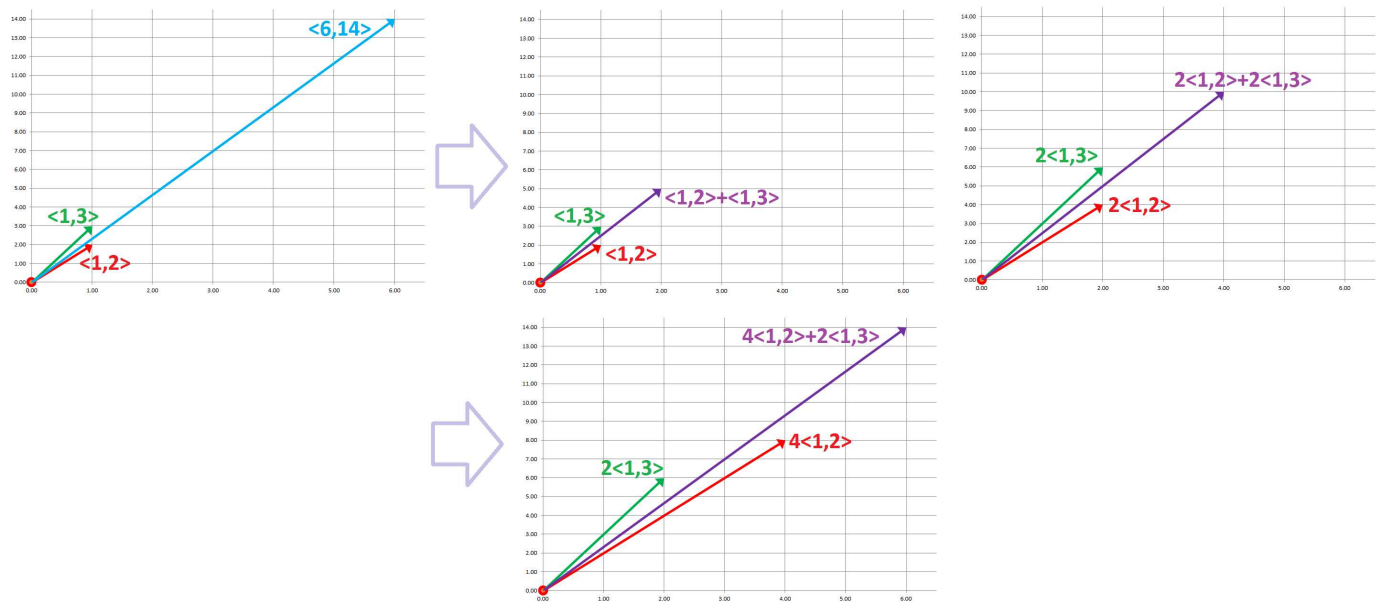
Prove the theorem.

Example 2.8.10: linear system

Recall from the beginning of the chapter how the solution to a system of linear equations was seen as if the two equations were equations about the coefficients, x and y , of *vectors* in the plane:

$$x \begin{bmatrix} 1 \\ 2 \end{bmatrix} + y \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 6 \\ 14 \end{bmatrix} .$$

To solve the system is to find a way to stretch these two vectors so that after adding them the result is the vector on the right:

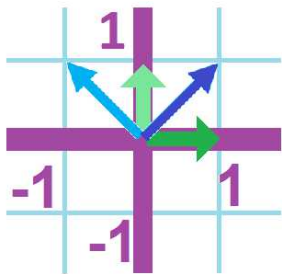


Now we conclude that we can guarantee that there is a solution only when the two vectors form a basis! For example, this mixture problem doesn't have a solution:

$$x \begin{bmatrix} 1 \\ 2 \end{bmatrix} + y \begin{bmatrix} 2 \\ 6 \end{bmatrix} = \begin{bmatrix} 6 \\ 14 \end{bmatrix} .$$

Example 2.8.11: coordinates

Consider an alternative basis along with the standard one:



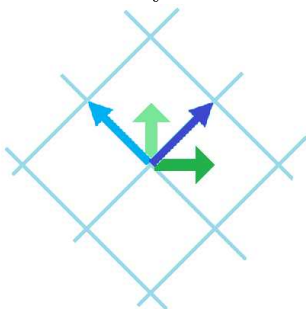
It's easy to express the new vectors in terms of old:

$$V_1 = \langle 1, 1 \rangle, \quad V_2 = \langle -1, 1 \rangle .$$

But what about vice versa? It's harder; we need to find a and b so that

$$e_1 = aV_1 + bV_2 ,$$

and the same for e_2 . We draw a grid for the new system to help:



The answer is:

$$e_1 = \frac{1}{2}V_1 - \frac{1}{2}V_2, \quad e_2 = \frac{1}{2}V_1 + \frac{1}{2}V_2 .$$

We can then re-write these vectors in the language of components *with respect to the new basis*:

$$e_1 = \left\langle \frac{1}{2}, -\frac{1}{2} \right\rangle = \frac{1}{2} \langle 1, -1 \rangle, \quad e_2 = \left\langle \frac{1}{2}, \frac{1}{2} \right\rangle = \frac{1}{2} \langle 1, 1 \rangle .$$

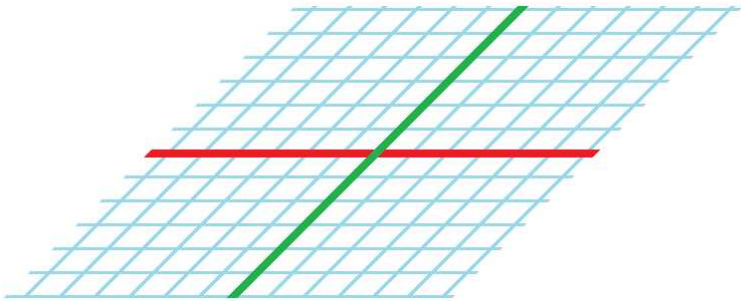
In summary, we have different coefficients and, therefore, different components of the same vectors with respect to different bases:

bases:	$\{e_1, e_2\}$	$\{V_1, V_2\}$
$V_1 =$	$\langle 1, 1 \rangle$	$= \langle 1, 0 \rangle$
$V_2 =$	$\langle -1, 1 \rangle$	$= \langle 0, 1 \rangle$

and

bases:	$\{e_1, e_2\}$	$\{V_1, V_2\}$
$e_1 =$	$\langle 1, 0 \rangle$	$= \left\langle \frac{1}{2}, -\frac{1}{2} \right\rangle$
$e_2 =$	$\langle 0, 1 \rangle$	$= \left\langle \frac{1}{2}, \frac{1}{2} \right\rangle$

In general, the vectors of the alternative basis might have any angle between them (as long as it's not zero). Then, we have a skewed grid:



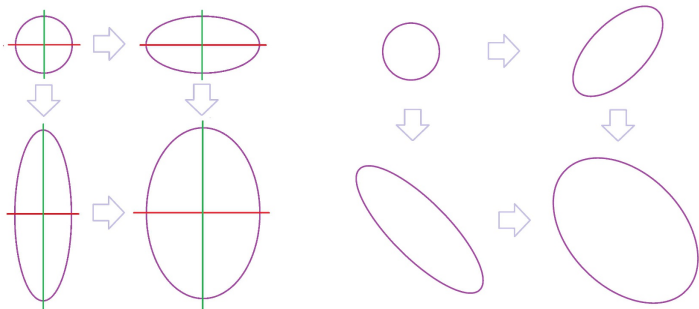
Thus, the component-wise algebra is fully operational whatever basis we choose:

$$\langle a, b \rangle + \langle c, d \rangle = \langle a + c, b + d \rangle \quad \text{and} \quad r \langle a, b \rangle = \langle ra, rb \rangle .$$

Warning!

Unlike the algebra, the geometry of the Cartesian system relies on the Pythagorean Theorem. As a result, the formulas for magnitudes and the dot products fail in the current form if used with non-perpendicular bases.

Now, the linear operators. They can be described *apart* from the Cartesian system (that was added to the vector space), i.e., rotation, stretching, etc.:



However, just as vector algebra works componentwise, so do linear operators. Furthermore, this approach works with respect to any basis:

Theorem 2.8.12: Linear Operator in Terms of Basis

Suppose $\{V_1, V_2\}$ is a basis. Then, all values of a linear operator $Y = F(X)$ are expressed as linear combinations of its values on these vectors; i.e., for any pair of real coefficients r_1 and r_2 , we have:

$$X = r_1V_1 + r_2V_2 \implies F(X) = r_1F(V_1) + r_2F(V_2).$$

Exercise 2.8.13

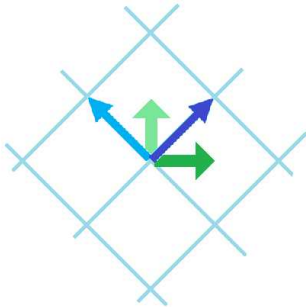
Prove the theorem.

In other words, the operator is fully determined by its values on the basis vectors – just as with the standard basis. But, just as with vectors, we have different matrices for the same linear operator with respect to different bases. The columns of the matrix of A are the values of the basis vectors under the operator:

$$A(V_1) \text{ and } A(V_2).$$

Example 2.8.14: matrices

We, again, consider this alternative basis along with the standard one:



The mutual representations are:

$$e_1 = \frac{1}{2}V_1 - \frac{1}{2}V_2, \quad e_2 = \frac{1}{2}V_1 + \frac{1}{2}V_2,$$

and

$$V_1 = e_1 + e_2, \quad V_2 = -e_1 + e_2.$$

Suppose an operator A stretches the plane along the x -axis (in other words, along e_1) by a factor of 2. Then the matrix of A with respect to $\{e_1, e_2\}$ is:

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}.$$

What about the other basis, $\{V_1, V_2\}$? It is hard to guess this time because – unlike for the standard basis – the change is not aligned with the basis vectors. Let’s use the “convenient” basis and then switch to the one we are interested in. We write the formulas in terms of e_1, e_2 first:

$$\begin{aligned} A(V_1) &= A(e_1 + e_2) &= A(e_1) + A(e_2) &= 2e_1 + e_2 \\ A(V_2) &= A(-e_1 + e_2) &= -A(e_1) + A(e_2) &= -2e_1 + e_2 \end{aligned}$$

Then we substitute the values of e_1, e_2 in terms of V_1, V_2 as written above:

$$\begin{aligned} A(V_1) &= 2e_1 + e_2 &= 2\left(\frac{1}{2}V_1 - \frac{1}{2}V_2\right) + \left(\frac{1}{2}V_1 + \frac{1}{2}V_2\right) &= \frac{3}{2}V_1 - \frac{1}{2}V_2 \\ A(V_2) &= -2e_1 + e_2 &= -2\left(\frac{1}{2}V_1 - \frac{1}{2}V_2\right) + \left(\frac{1}{2}V_1 + \frac{1}{2}V_2\right) &= -\frac{1}{2}V_1 + \frac{3}{2}V_2 \end{aligned}$$

Therefore, the matrix of the linear operator with respect to $\{V_1, V_2\}$ is:

$$A = \begin{bmatrix} \frac{3}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{3}{2} \end{bmatrix}.$$

What if the stretch was along V_1 ? This operator's matrix with respect to $\{V_1, V_2\}$ is simple:

$$B = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}.$$

Exercise 2.8.15

Rewrite the above computation in terms of components.

2.9. Classification of linear operators according to their eigenvalues

We apply the last theorem to eigenvectors.

Corollary 2.9.1: Representation in Terms of Eigenvectors

Suppose V_1 and V_2 are two eigenvectors of a linear operator F that correspond to two (possibly equal) eigenvalues λ_1 and λ_2 . Suppose also that V_1 and V_2 aren't multiples of each other. Then, all values of the linear operator $Y = F(X)$ are represented as linear combinations of its values on the eigenvectors:

$$X = r_1V_1 + r_2V_2 \implies F(X) = r_1\lambda_1V_1 + r_2\lambda_2V_2,$$

with some real coefficients r_1 and r_2 .

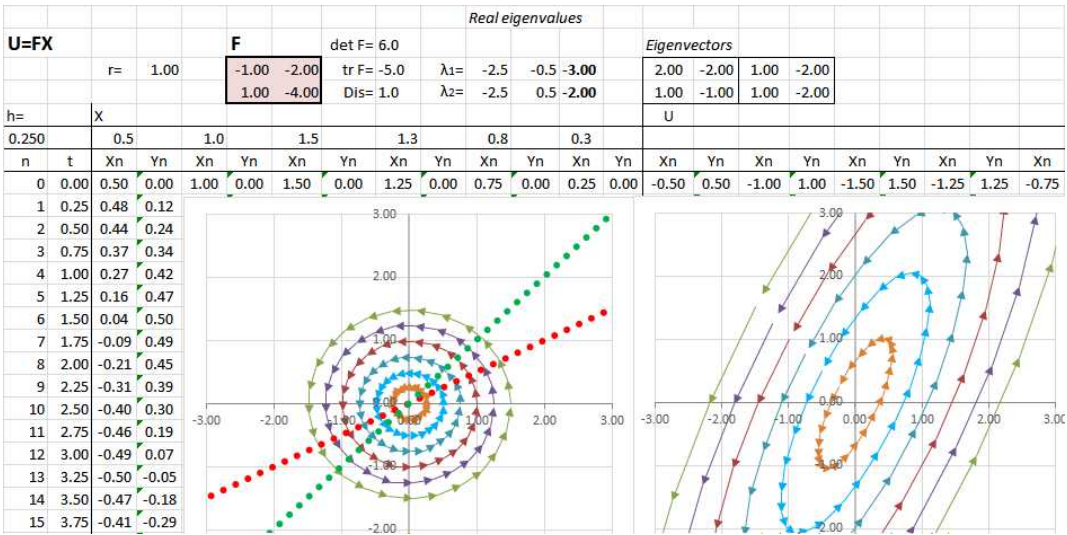
In other words, the matrix of F with respect to the basis $\{V_1, V_2\}$ of eigenvectors is diagonal with the eigenvalues on the diagonal:

$$F = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}.$$

Example 2.9.2: stretch-shrink

Let's consider this function:

$$\begin{cases} u &= -x & - & 2y \\ v &= x & - & 4y \end{cases} \implies F = \begin{bmatrix} -1 & -2 \\ 1 & -4 \end{bmatrix}.$$



Let’s confirm what is shown above. The analysis starts with the characteristic polynomial:

$$\det(F - \lambda I) = \det \begin{bmatrix} -1 - \lambda & -2 \\ 1 & -4 - \lambda \end{bmatrix} = \lambda^2 - 5\lambda + 6.$$

Therefore, the *eigenvalues* are:

$$\lambda_1 = -3, \lambda_2 = -2.$$

To find the *eigenvectors*, we solve the two vector equations:

$$FV_i = \lambda_i V_i, \quad i = 1, 2.$$

The first, $\lambda_1 = -3$:

$$FV_1 = \begin{bmatrix} -1 & -2 \\ 1 & -4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = -3 \begin{bmatrix} x \\ y \end{bmatrix}.$$

This gives us the following system of linear equations:

$$\begin{cases} -x - 2y = -3x \\ x - 4y = -3y \end{cases} \implies \begin{cases} 2x - 2y = 0 \\ x - y = 0 \end{cases} \implies x = y.$$

We have discovered, again, that this is the same equation; this line gives us the eigenspace. We choose one eigenvector:

$$V_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

The second eigenvalue:

$$FV_2 = \begin{bmatrix} -1 & -2 \\ 1 & -4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = -2 \begin{bmatrix} x \\ y \end{bmatrix}.$$

We have the following system (same equation):

$$\begin{cases} -x - 2y = -2x \\ x - 4y = -2y \end{cases} \implies \begin{cases} x - 2y = 0 \\ x - 2y = 0 \end{cases} \implies x = 2y.$$

This line is the eigenspace of $\lambda_2 = -2$. We choose one eigenvector:

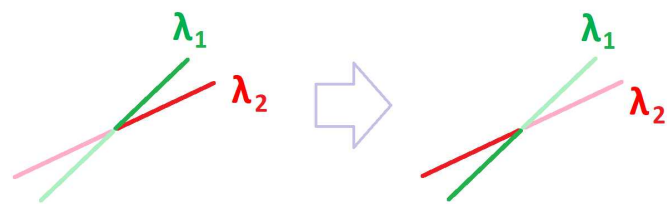
$$V_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

The pair $\{V_1, V_2\}$ is a basis!

We summarize what F does:

1. A flip and stretch along the vector $\langle 1, 1 \rangle$: The line $y = x$ remains intact.
2. A flip and stretch along the vector $\langle 2, 1 \rangle$: The line $x = 2y$ remains intact.

We also conclude that there is no change of orientation. Stretching aside, this looks like central symmetry:



We observe fanning between these two lines. For the rest of the vectors, we have:

$$X = xV_1 + yV_2 \implies FX = -3x \begin{bmatrix} 2 \\ -1 \end{bmatrix} - 2y \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

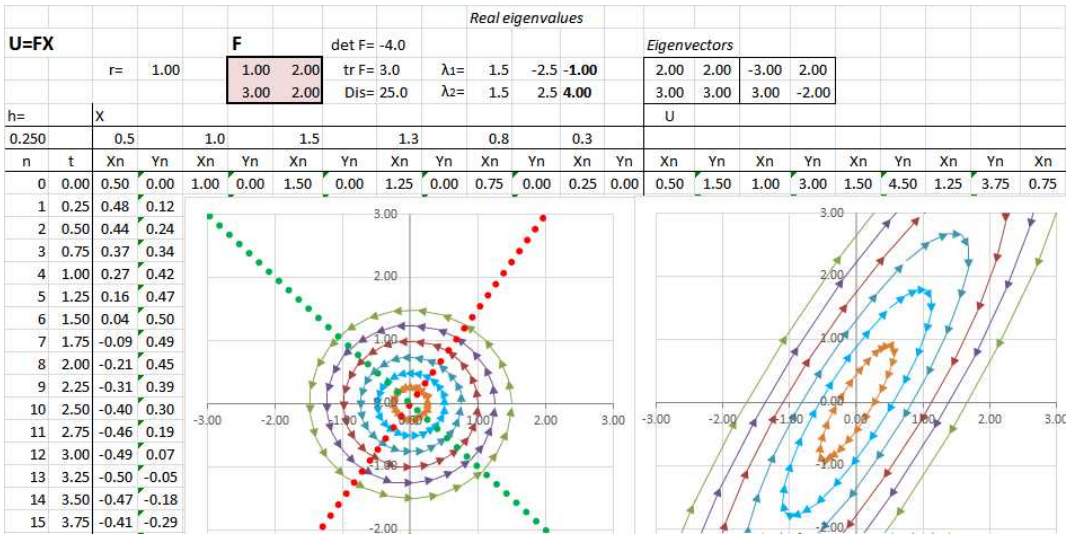
Therefore, matrix of F with respect to the basis $\{V_1, V_2\}$ is diagonal:

$$F = \begin{bmatrix} -3 & 0 \\ 0 & -2 \end{bmatrix}.$$

Example 2.9.3: stretch-shrink

Let’s consider this linear operator:

$$\begin{cases} u = x + 2y \\ v = 3x + 2y \end{cases} \implies F = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}.$$



Let’s find the eigenvectors:

$$\det(F - \lambda I) = \det \begin{bmatrix} 1 - \lambda & 2 \\ 3 & 2 - \lambda \end{bmatrix} = \lambda^2 - 3\lambda - 4.$$

Therefore, the eigenvalues are:

$$\lambda_1 = -1, \lambda_2 = 4.$$

Now we find the eigenvectors. We solve the two equations:

$$FV_i = \lambda_i V_i, \quad i = 1, 2.$$

The first:

$$FV_1 = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = -1 \begin{bmatrix} x \\ y \end{bmatrix}.$$

This gives as the following system of linear equations:

$$\begin{cases} x + 2y = -x \\ 3x + 2y = -y \end{cases} \implies \begin{cases} 2x + 2y = 0 \\ 3x + 3y = 0 \end{cases} \implies x = -y.$$

We choose:

$$V_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} .$$

Every value within this eigenspace (the line $y = -x$) is a multiple of this eigenvector:

$$X = \lambda_1 V_1 = - \begin{bmatrix} 1 \\ -1 \end{bmatrix} .$$

The second eigenvalue:

$$FV_2 = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 4 \begin{bmatrix} x \\ y \end{bmatrix} .$$

We have the following system:

$$\begin{cases} x + 2y = 4x \\ 3x + 2y = 4y \end{cases} \implies \begin{cases} -3x + 2y = 0 \\ 3x - 2y = 0 \end{cases} \implies x = 2y/3 .$$

We choose:

$$V_2 = \begin{bmatrix} 1 \\ 3/2 \end{bmatrix} .$$

Every value within this eigenspace (the line $y = 3x/2$) is a multiple of this eigenvector:

$$X = \lambda_2 V_2 = 4 \begin{bmatrix} 1 \\ 3/2 \end{bmatrix} .$$

The pair $\{V_1, V_2\}$ is a basis! Then,

$$X = xV_1 + yV_2 \implies U = F(X) = -xV_1 + 4yV_2 .$$

The matrix of F with respect to the basis $\{V_1, V_2\}$ is:

$$F = \begin{bmatrix} -1 & 0 \\ 0 & 4 \end{bmatrix} .$$

We observe fanning between these two lines.

Let’s summarize the results.

Theorem 2.9.4: Classification of Linear Operators – Real Eigenvalues

Suppose matrix F has two real non-zero eigenvalues λ_1 and λ_2 . Then, the function $U = F(X)$ stretches/shrinks the two eigenspace by factors $|\lambda_1|$ and $|\lambda_2|$ respectively and, furthermore:

- If λ_1 and λ_2 have the same sign, it preserves the orientation of a closed curve around the origin.
- If λ_1 and λ_2 have the opposite signs, it reverses the orientation of a closed curve around the origin.

Exercise 2.9.5

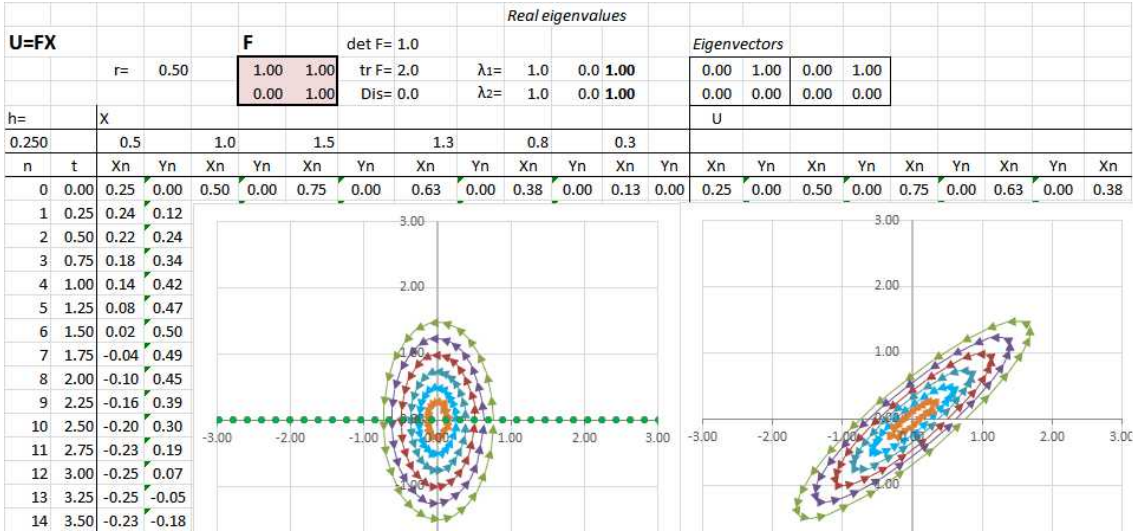
Apply the theorem to the last example.

Example 2.9.6: skewing-shearing

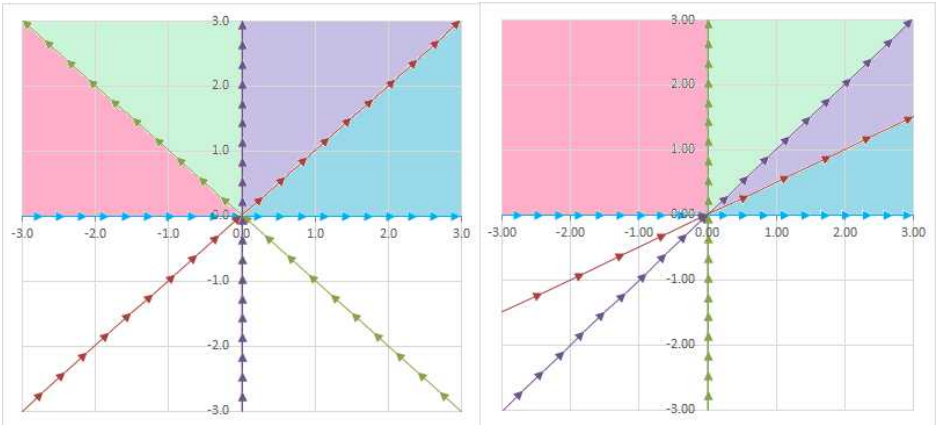
Consider this matrix:

$$F = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} .$$

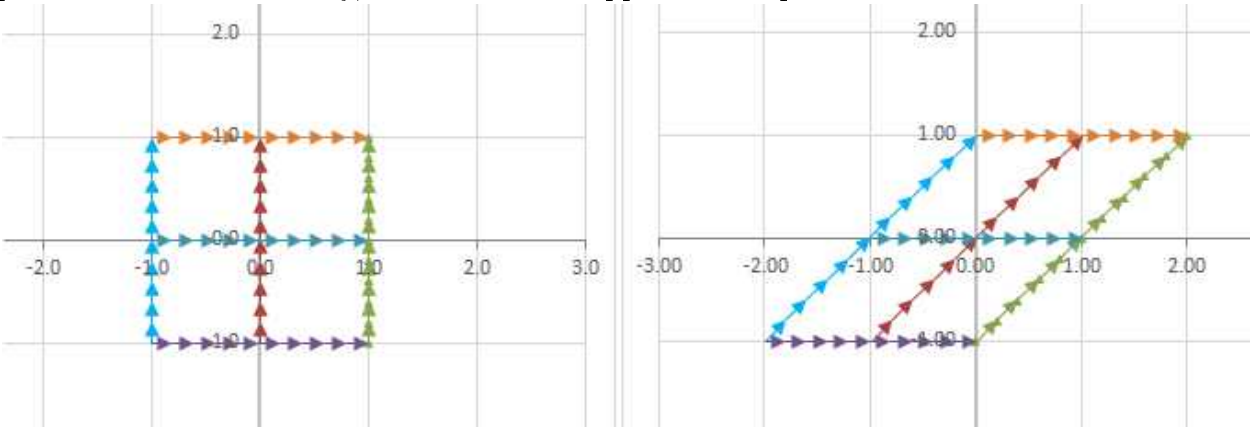
Below, we replace a circle with an ellipse to see what happens to it under such a function:



There is still angular stretch-shrink but this time it is between the two ends of the same line. We see “fanning out” again:



This time, however, the fan is fully open! It makes a difference that the fanning happens to a whole half-plane. To see more clearly, consider what happens to a square:



This is the characteristic polynomial:

$$\det(F - \lambda I) = \det \begin{bmatrix} 1 - \lambda & 1 \\ 0 & 1 - \lambda \end{bmatrix} = (1 - \lambda)^2.$$

Therefore, the eigenvalues are

$$\lambda_1 = \lambda_2 = 1.$$

What are the eigenvectors?

$$FV = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 1 \begin{bmatrix} x \\ y \end{bmatrix}.$$

This gives as the following system of linear equations:

$$\begin{cases} x + y = x \text{ AND} \\ y = y \end{cases} \implies x \text{ any, } y = 0.$$

The only eigenvectors are horizontal! Therefore, our classification theorem doesn't apply. There is no diagonal matrix for this operator.

Example 2.9.7: rotations

There are other outcomes that the theorem doesn't cover. Recall the characteristic polynomial of the matrix A of the 90-degree rotation:

$$\chi_A(\lambda) = \det \begin{bmatrix} -\lambda & -1 \\ 1 & -\lambda \end{bmatrix} = \lambda^2 + 1.$$

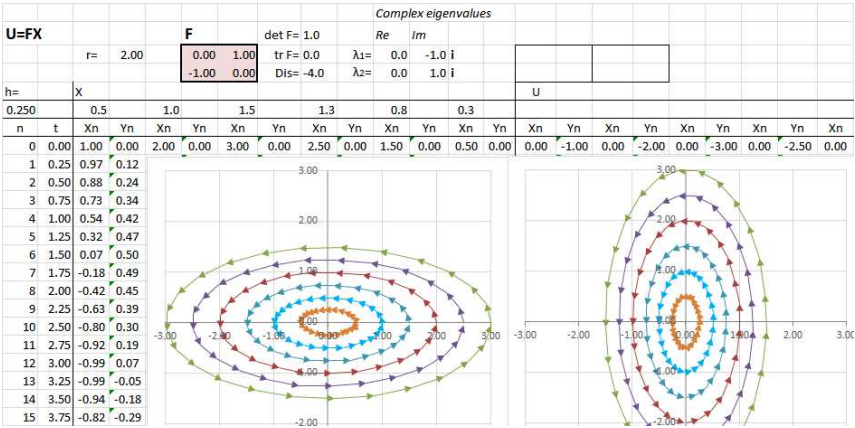
But the characteristic equation,

$$x^2 + 1 = 0,$$

has no solutions! Are we done then? Not if we are willing to use *complex numbers* (next chapter):

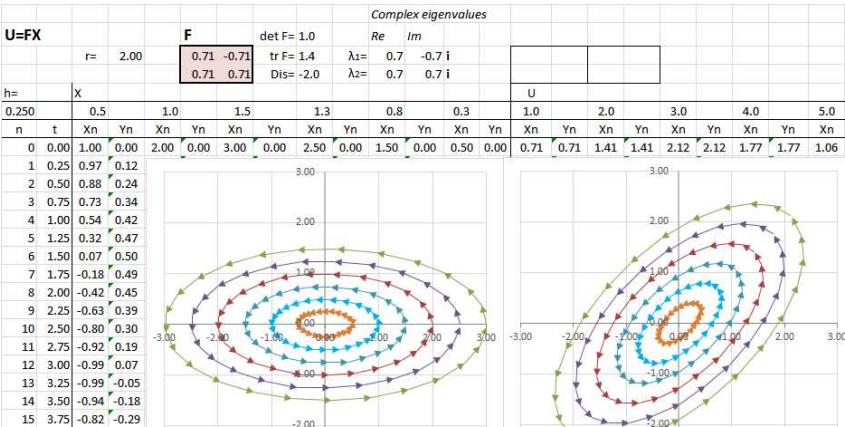
$$\lambda_{1,2} = i \text{ and } -i.$$

This is the effect of rotation:



Let's consider a rotation through an arbitrary angle θ :

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$



The angle can be seen in the characteristic polynomial:

$$\chi_A(\lambda) = (\cos \theta - \lambda)^2 + \sin^2 \theta = \cos^2 \theta - 2 \cos \theta \lambda + \lambda^2 + \sin^2 \theta = \lambda^2 - 2 \cos \theta \lambda + 1.$$

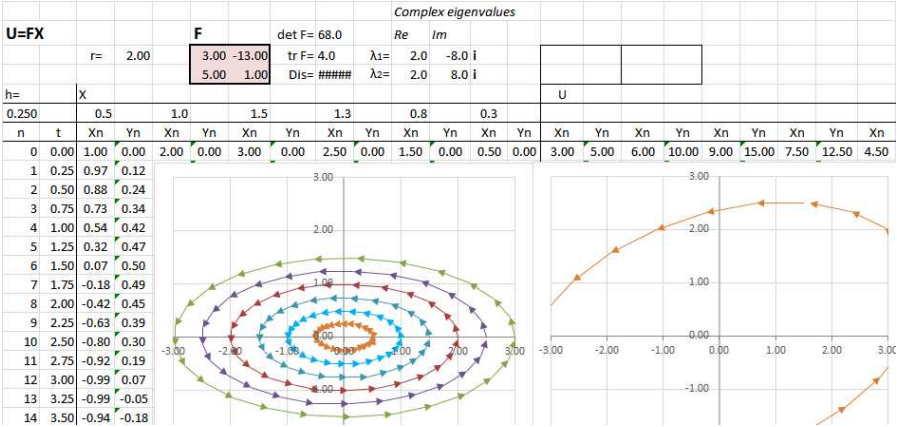
The discriminant of this polynomial is negative. Therefore, it has no real roots. The result makes sense: A rotation cannot possibly have eigenvectors because all vectors are rotated!

Example 2.9.8: rotation with stretch-shrink

Let’s consider this linear operator:

$$\begin{cases} u &= 3x - 13y, \\ v &= 5x + y, \end{cases} \quad \text{and } F = \begin{bmatrix} 3 & -13 \\ 5 & 1 \end{bmatrix}.$$

Below we can recognize both rotation and re-scaling:



This is our characteristic polynomial:

$$\chi(\lambda) = \det(F - \lambda I) = \det \begin{bmatrix} 3 - \lambda & -13 \\ 5 & 1 - \lambda \end{bmatrix} = \lambda^2 - 4\lambda + 68.$$

Exercise 2.9.9

What does it tell us?

Our interpretation of the characteristic polynomial in terms of the trace of the matrix:

$$\chi(\lambda) = \lambda^2 - \text{tr } F \lambda + \det F.$$

allows us to prove in the next chapter the following result for the case of no real eigenvalues:

Corollary 2.9.10: Trace and Discriminant

Suppose the discriminant of the characteristic polynomial of a matrix F satisfies:

$$D = (\text{tr } F)^2 - 4 \det F \leq 0.$$

Then, the operator $U = FX$ does the following:

1. It rotates the real plane through the following angle:

$$\theta = \sin^{-1} \left(\frac{1}{2} \sqrt{\frac{4 - (\text{tr } F)^2}{\det F}} \right).$$

2. It re-scales the plane uniformly by the following factor:

$$s = \sqrt{\det F}.$$

Exercise 2.9.11

Apply the corollary to the last example.

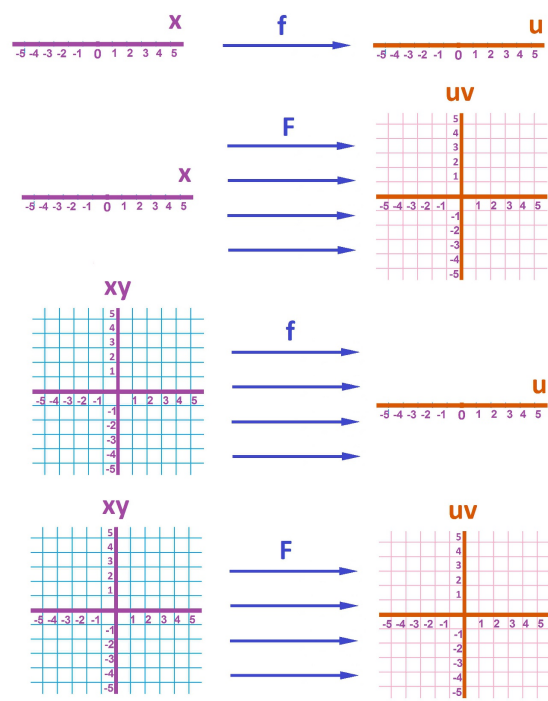
Chapter 3: Vector and complex variables

Contents

3.1 Algebra of linear operators and matrices	198
3.2 Compositions of linear operators	204
3.3 How complex numbers emerge	209
3.4 Classification of quadratic polynomials	216
3.5 The complex plane \mathbf{C} is the Euclidean space \mathbf{R}^2	219
3.6 Multiplication of complex numbers: \mathbf{C} isn't just \mathbf{R}^2	223
3.7 Complex functions	228
3.8 Complex linear operators	233
3.9 Linear operators with complex eigenvalues	236
3.10 Complex calculus	242
3.11 Series and power series	244
3.12 Solving ODEs with power series	250

3.1. Algebra of linear operators and matrices

We will take a broader view at linear operator and include the lower dimensions. These are the four possibilities:



We want to understand how these operators are represented by matrices and how these matrices are combined to produce *compositions*.

The rule remains:

- The value under F of each basis vector in the domain of F becomes a column in the matrix of F .

Let’s apply the rule to these four situations using the standard bases.

Example 3.1.1: dimensions 1 and 1

Suppose we have a linear operator, which is just a numerical function:

$$f : \mathbf{R} \rightarrow \mathbf{R} \text{ defined by } f(x) = 3x .$$

It is a stretch by a factor of 3. What is its matrix? The basis of the x -axis is $\langle 1 \rangle$ and the basis of the u -axis is $\langle 1 \rangle$. The operator works as follows:

$$f(\langle 1 \rangle) = 3 \langle 1 \rangle .$$

Therefore, its matrix is

$$f = [3] .$$

Example 3.1.2: dimensions 1 and 2

Suppose we have a linear operator, which is just a parametric curve:

$$F : \mathbf{R} \rightarrow \mathbf{R}^2 \text{ defined by } F(x) = \langle 3x, 2x \rangle .$$

It stretches the x -axis on the uv -plane along the vector $\langle 3, 2 \rangle$. What is its matrix? The basis of the x -axis is $\langle 1 \rangle$ and the basis of the uv -plane is $\langle 1, 0 \rangle$ and $\langle 0, 1 \rangle$. The operator works as follows:

$$F(\langle 1 \rangle) = \langle 3, 2 \rangle .$$

Therefore, its matrix is

$$F = \begin{bmatrix} 3 \\ 2 \end{bmatrix} .$$

Example 3.1.3: dimensions 2 and 1

Suppose we have a linear operator, which is just a function of two variables:

$$f : \mathbf{R}^2 \rightarrow \mathbf{R} \text{ defined by } f(x, y) = 3x + 2y .$$

It rolls the xy -plane on the u -axis. What is its matrix? The basis of the xy -plane is $\langle 1, 0 \rangle$ and $\langle 0, 1 \rangle$ and the basis of the u -axis is $\langle 1 \rangle$. The operator works as follows:

$$f(\langle 1, 0 \rangle) = \langle 3 \rangle \quad \text{and} \quad f(\langle 0, 1 \rangle) = \langle 2 \rangle .$$

Therefore, its matrix is

$$f = [3 \ 2] .$$

Example 3.1.4: dimensions 2 and 2

Suppose we have a linear operator, which is just a transformation of the plane:

$$F : \mathbf{R}^2 \rightarrow \mathbf{R}^2 \text{ defined by } f(x, y) = \langle 3x + 2y, 5x - y \rangle .$$

We’d need the eigenvector analysis in order to determine what it does... What is its matrix? The

basis of the xy -plane is $\langle 1, 0 \rangle$ and $\langle 0, 1 \rangle$ and the basis of the u -axis is $\langle 1, 0 \rangle$ and $\langle 0, 1 \rangle$. The operator works as follows:

$$F(\langle 1, 0 \rangle) = \langle 3, 5 \rangle \quad \text{and} \quad F(\langle 0, 1 \rangle) = \langle 2, -1 \rangle .$$

Therefore, its matrix is

$$F = \begin{bmatrix} 3 & 2 \\ 5 & -1 \end{bmatrix} .$$

As you can see, we can jump ahead of the rule described above and write the coefficients present in the formula of the linear operator straight into the matrix.

Warning!

Its matrix is just an abbreviated representation of a linear operator.

Exercise 3.1.5

Include the dimension 0 for domains and codomains in the above analysis.

Whenever there is algebra in the output space, we can use it to do algebra of the functions. If the codomain of functions is a vector space, we can add these functions and multiply them by a constant. We just narrow down this idea to linear operators:

Definition 3.1.6: addition of linear operators

Given two linear operators:

$$F, G : \mathbf{R}^n \rightarrow \mathbf{R}^m ,$$

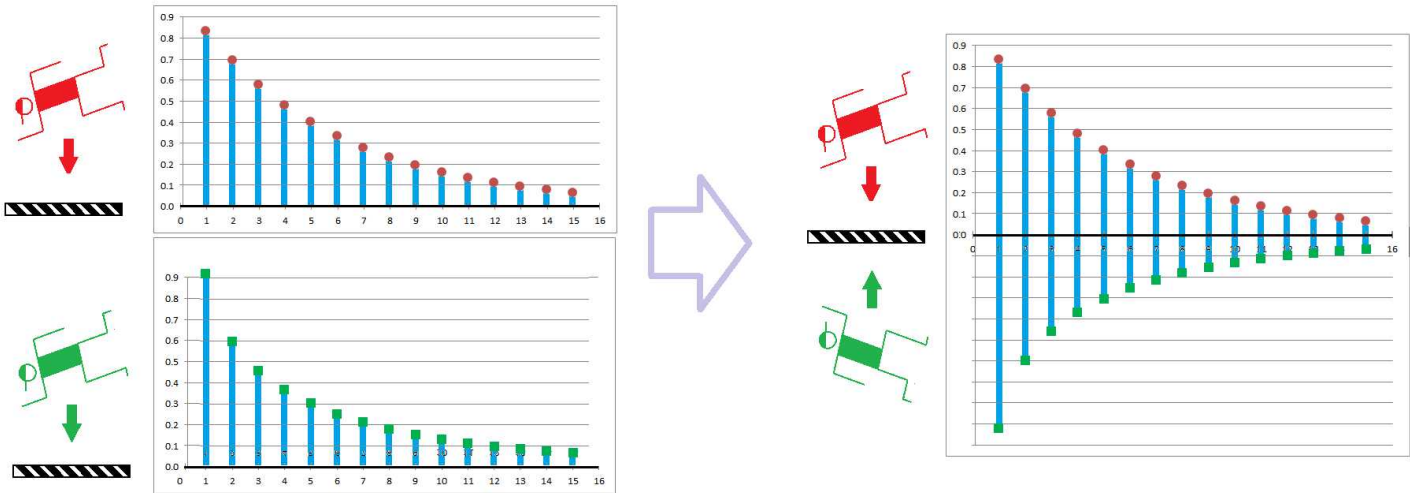
their *sum* is linear operator:

$$F + G : \mathbf{R}^n \rightarrow \mathbf{R}^m ,$$

defined by:

$$(F + G)(x) = F(x) + G(x) .$$

We illustrate this operation just as before:



But if these are linear operators, what happens to their matrices? We go though the same four cases below.

Example 3.1.7: dimensions 1 and 1

Given linear operators (numerical functions):

$$f, g : \mathbf{R} \rightarrow \mathbf{R} \text{ defined by } f(x) = 3x \text{ and } g(x) = 2x .$$

Their matrices are:

$$f = [3] \text{ and } g = [2] .$$

What about their sum? It is an operator with the same domain and codomain and it is computed as follows:

$$f + g : \mathbf{R} \rightarrow \mathbf{R} \text{ defined by } (f + g)(x) = f(x) + g(x) = 3x + 2x = 5x .$$

Its matrix is

$$f + g = [5] .$$

Of course, this new number is just the sum of the two original numbers.

Example 3.1.8: dimensions 1 and 2

Given linear operators (parametric curves):

$$F, G : \mathbf{R} \rightarrow \mathbf{R}^2 \text{ defined by } F(x) = \langle 3x, 2x \rangle \text{ and } G(x) = \langle 5x, -x \rangle .$$

Their matrices are:

$$F = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \text{ and } G = \begin{bmatrix} 5 \\ -1 \end{bmatrix} .$$

What about their sum? It is an operator with the same domain and codomain and it is computed as follows:

$$F + G : \mathbf{R} \rightarrow \mathbf{R}^2 \text{ defined by } (F + G)(x) = F(x) + G(x) = \langle 3x, 2x \rangle + \langle 5x, -x \rangle = \langle 8x, x \rangle .$$

Its matrix is

$$F + G = \begin{bmatrix} 8 \\ 1 \end{bmatrix} .$$

Of course, this is just the sum of the two as if they were vectors.

Example 3.1.9: dimensions 2 and 1

Given linear operators (functions of two variables):

$$f, g : \mathbf{R}^2 \rightarrow \mathbf{R} \text{ defined by } f(x, y) = 3x + 2y \text{ and } g(x, y) = 5x - 2y .$$

Their matrices are:

$$f = [3, 2] \text{ and } g = [5, -2] .$$

Their sum is an operator with the same domain and codomain and it is computed as follows (this is vector addition):

$$f + g : \mathbf{R}^2 \rightarrow \mathbf{R} \text{ defined by } (f + g)(x, y) = f(x, y) + g(x, y) = 3x + 2y + 5x - 2y = 8x .$$

Its matrix is

$$f = [8, 0] ,$$

the sum – componentwise – of the two.

Example 3.1.10: dimensions 2 and 2

Given linear operators (transformations of the plane):

$F, G : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ defined by $F(x, y) = \langle 3x + 2y, 5x - y \rangle$ and $G(x, y) = \langle 5x + y, x + y \rangle$.

Their matrices are:

$$F = \begin{bmatrix} 3 & 2 \\ 5 & -1 \end{bmatrix} \quad \text{and} \quad G = \begin{bmatrix} 5 & 1 \\ 1 & 1 \end{bmatrix}.$$

Their sum is an operator with the same domain and codomain and it is computed as follows (this is vector addition):

$F + G : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ defined by $(F + G)(x, y) = \langle 3x + 2y, 5x - y \rangle + \langle 5x + y, x + y \rangle = \langle 8x + 3y, 6x \rangle$.

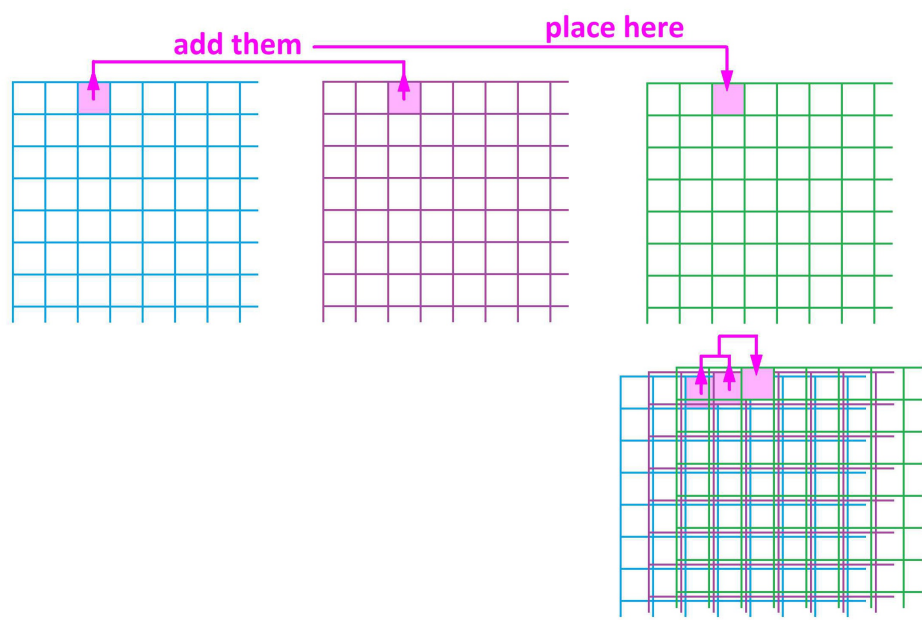
Its matrix is

$$F + G = \begin{bmatrix} 8 & 3 \\ 6 & 0 \end{bmatrix},$$

the sum – componentwise – of the two:

$$F + G = \begin{bmatrix} 3 & 2 \\ 5 & -1 \end{bmatrix} + \begin{bmatrix} 5 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 + 5 & 2 + 1 \\ 5 + 1 & (-1) + 1 \end{bmatrix} = \begin{bmatrix} 8 & 3 \\ 6 & 0 \end{bmatrix}.$$

This operation of matrix addition is componentwise: The two operators have the same domain and codomain, so that the dimensions of the matrices are equal too. They can then be overlapped so that the entries are aligned and added accordingly:



Definition 3.1.11: addition of matrices

Suppose A and B are two $m \times n$ matrices. Then their *sum* is the $m \times n$ matrix, denoted by:

$A + B$

the ij -entry of which is the sum of the ij -entries of A and B .

In other words, if $A = a_{ij}$, $B = b_{ij}$, and $C = A + B = c_{ij}$, then

$$c_{ij} = a_{ij} + b_{ij},$$

for each $i = 1, 2, \dots, m$ and each $j = 1, 2, \dots, n$.

It is simpler for scalar multiplication:

Definition 3.1.12: scalar multiplication of linear operator

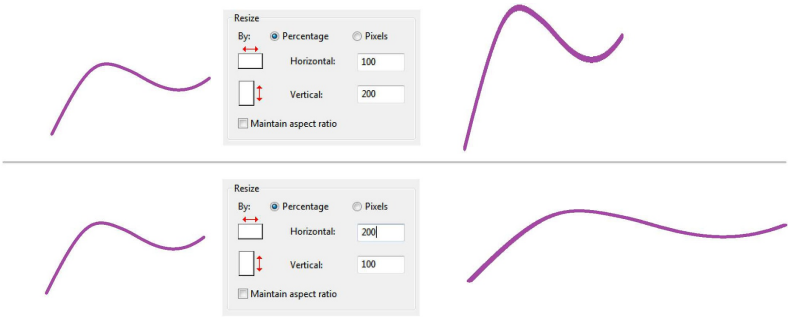
Given a linear operator:

$$F : \mathbf{R}^n \rightarrow \mathbf{R}^m,$$

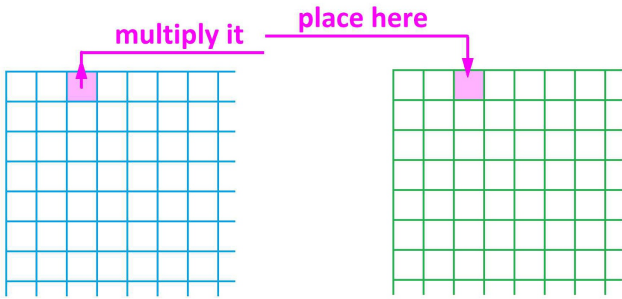
its *scalar product* with a real number r is linear operator:

$$rF : \mathbf{R}^n \rightarrow \mathbf{R}^m,$$

defined by:

$$(rF)(x) = rF(x).$$


This operation is component-wise again: Every entry is multiplied by the same number.



Definition 3.1.13: scalar multiplication of matrices

Suppose A is an $m \times n$ matrix. Then its *scalar multiple* by a real number r , denoted by:

$$rA$$

is an $m \times n$ matrix, the ij -entry of which is the product of the ij -entry of A by r .

In other words, if $A = a_{ij}$ and $C = rA = c_{ij}$, then

$$c_{ij} = ra_{ij},$$

for each $i = 1, 2, \dots, m$ and each $j = 1, 2, \dots, n$.

Warning!

The matrix is just an abbreviated representation of a linear operator. Accordingly, the matrix operations are just abbreviated representations of the operations on linear operators.

3.2. Compositions of linear operators

Let’s take the problem about *mixtures* to the next level.

We have:

1. n ingredients and, therefore, n unknowns or variables x_1, \dots, x_n representing the amounts of each; and
2. m requirements or restrictions, i.e., m linear equations involving these variables ($k = 1, 2, \dots, m$):

$$a_{k1}x_1 + \dots + a_{kn}x_n = b_k .$$

This system of linear equations is very cumbersome.

As before, we translate this a system into a vector-matrix equation:

$F X = B$

where

1. $X = \langle x_1, \dots, x_n \rangle$ is the vector of the unknowns,
2. $B = \langle b_1, \dots, b_m \rangle$ is the vector of the totals, and
3. $F = a_{ij}$ is the $m \times n$ matrix made up of the coefficients of the system of linear equations.

In light of the recent development, we prefer to look at the equation as the following:

$F(X) = B$

i.e., we have a linear operator:

$$F : \mathbf{R}^n \rightarrow \mathbf{R}^m .$$

And the equation needs to be solved!

In dimension 1, the equation $kx = b$ is solved by undoing the multiplication by k by division by k :

$$kx = b \implies x = \frac{b}{k} .$$

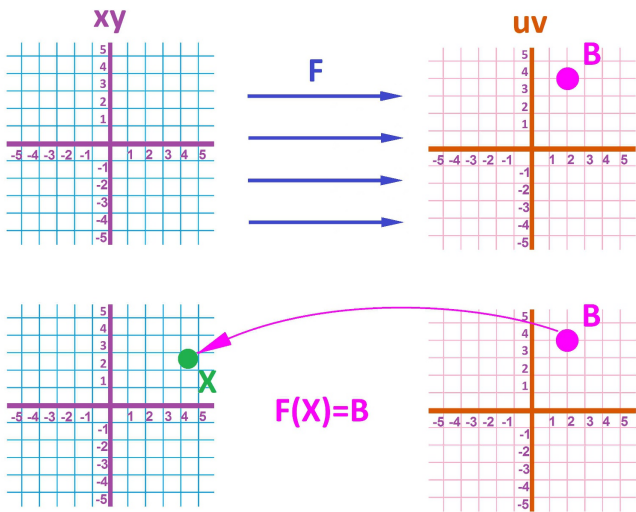
Similarly, we need the *inverse*

$$F^{-1} : \mathbf{R}^m \rightarrow \mathbf{R}^n$$

of F to solve our equation:

$F(X) = B \implies X = F^{-1}(B)$

As an illustration, the operator F transforms the xy -plane into the, say, uv -plane:



One particular vector in the uv -plane, B , needs to be traced back to the xy -plane. Of course, if we have F^{-1} , we'll find the counterparts for all B 's.

Example 3.2.1: transformations of the plane

We know some of the answers:

1. If F is the uniform stretch by 2, then F^{-1} is the uniform shrink by 2.
2. If F is the stretch by 2 in the direction of a vector $e_1 = \langle 1, 0 \rangle$, then F^{-1} is the uniform shrink by 2 in the direction of e_1 .
3. If F is the rotation by 90 degrees clockwise, then F^{-1} is the rotation by 90 degrees counterclockwise.
4. If F is the flip about the x -axis, then F^{-1} is the flip about the x -axis.

In other words,

1.

$F(X) = 2X$

$\implies F^{-1}(Y) = \frac{1}{2}Y$

2.

$F = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$

$\implies F^{-1} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix}$

3.

$F = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$

$\implies F^{-1} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$

4.

$F = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$

$\implies F^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$

Let's recall that the inverse is defined via compositions. It must satisfy:

$$F(F^{-1}(Y)) = Y,$$

for all Y , and

$$F^{-1}(F(X)) = X,$$

for all X . In other words,

$$F \circ F^{-1} = I,$$

and

$$F^{-1} \circ F = I,$$

where I is the identity matrix.

We need to understand compositions better.

We know how to compute compositions of functions. This is the composition:

\mathbf{R}^n

\xrightarrow{F}

\mathbf{R}^m

\xrightarrow{G}

\mathbf{R}^k

It is computed via substitution:

$$(G \circ F)(X) = G(F(X)).$$

But what happens to the matrices? How are the matrices of F and G combined to produce the matrix of $G \circ F$? It is called *matrix multiplication*. Here is why.

Example 3.2.2: compositions $\mathbf{R}^1 \rightarrow \mathbf{R}^1 \rightarrow \mathbf{R}^1$

Given linear operators (numerical functions):

$$f : \mathbf{R} \rightarrow \mathbf{R} \text{ defined by } f(x) = 3x,$$

and

$$g : \mathbf{R} \rightarrow \mathbf{R} \text{ defined by } g(y) = 2y.$$

Their matrices are:

$$f = [3] \quad \text{and} \quad g = [2] .$$

What about their composition? The codomain of the former and the domain of the latter match! The composition is computed as follows:

$$g \circ f : \mathbf{R} \rightarrow \mathbf{R} \quad \text{defined by} \quad (g \circ f)(x) = g(f(x)) = 2(3x) = 6x .$$

Its matrix is

$$g \circ f = [6] .$$

Of course, this new number is just the *product* of the two original numbers.

Example 3.2.3: compositions $\mathbf{R}^1 \rightarrow \mathbf{R}^1 \rightarrow \mathbf{R}^2$

Given linear operators (a numerical function and a parametric curve):

$$f : \mathbf{R} \rightarrow \mathbf{R} \quad \text{defined by} \quad f(x) = 3x ,$$

and

$$G : \mathbf{R} \rightarrow \mathbf{R}^2 \quad \text{defined by} \quad G(y) = \langle 3y, 2y \rangle .$$

Their matrices are:

$$f = [3] \quad \text{and} \quad G = \begin{bmatrix} 3 \\ 2 \end{bmatrix} .$$

What about their composition? It is an operator (a parametric curve) computed as follows:

$$G \circ f : \mathbf{R} \rightarrow \mathbf{R}^2 ,$$

defined by

$$(G \circ f)(x) = G(f(x)) = \langle 3(3x), 2(3x) \rangle = \langle 9x, 6x \rangle .$$

Its matrix is

$$G \circ f = \begin{bmatrix} 9 \\ 6 \end{bmatrix} .$$

Of course, this is just the product of the two as if the first is a number and the second a vector:

$$Gf = \begin{bmatrix} 3 \\ 2 \end{bmatrix} [3] = \begin{bmatrix} 3 \cdot 3 \\ 2 \cdot 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 6 \end{bmatrix} .$$

Example 3.2.4: compositions $\mathbf{R}^1 \rightarrow \mathbf{R}^2 \rightarrow \mathbf{R}^1$

Given linear operators (a parametric curve and a function of two variables):

$$F : \mathbf{R} \rightarrow \mathbf{R}^2 \quad \text{defined by} \quad F(x) = \langle 3x, 2x \rangle ,$$

and

$$g : \mathbf{R}^2 \rightarrow \mathbf{R} \quad \text{defined by} \quad g(u, v) = 5u - 2v .$$

Their matrices are:

$$F = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \quad \text{and} \quad g = [5, -2] .$$

Their sum is a numerical function and it is computed as follows:

$$g \circ F : \mathbf{R} \rightarrow \mathbf{R} ,$$

defined by

$$(g \circ F)(x) = g(F(x)) = 5(3x) - 2(2x) = 11x .$$

Its matrix is

$$g \circ F = [11] .$$

It's the dot product of the two vector-like matrices:

$$gF = [5, -2] \begin{bmatrix} 3 \\ 2 \end{bmatrix} = [5 \cdot 3 + (-2) \cdot 2] = [11] .$$

Example 3.2.5: compositions $\mathbf{R}^2 \rightarrow \mathbf{R}^2 \rightarrow \mathbf{R}^2$

Given linear operators (transformations of the planes):

$$F, G : \mathbf{R}^2 \rightarrow \mathbf{R}^2$$

defined by

$$F(x, y) = \langle u, v \rangle = \langle 3x + 2y, 5x - y \rangle \quad \text{and} \quad G(u, v) = \langle 5u + v, u + v \rangle .$$

Their matrices are:

$$F = \begin{bmatrix} 3 & 2 \\ 5 & -1 \end{bmatrix} \quad \text{and} \quad G = \begin{bmatrix} 5 & 1 \\ 1 & 1 \end{bmatrix} .$$

Their composition is an operator computed by substitution:

$$G \circ F : \mathbf{R}^2 \rightarrow \mathbf{R}^2 ,$$

defined by

$$(G \circ F)(x, y) = \langle 5(3x + 2y) + (5x - y), (3x + 2y) + (5x - y) \rangle = \langle 20x + 9y, 8x + y \rangle .$$

Its matrix is

$$G \circ F = \begin{bmatrix} 20 & 9 \\ 8 & 1 \end{bmatrix} .$$

It is seen as computed via four dot products of the four pairs of rows (from the first matrix) and columns (from the second):

$\begin{bmatrix} 5 & 1 \\ 1 & 1 \end{bmatrix}$

\cdot

$\begin{bmatrix} 3 & 2 \\ 5 & -1 \end{bmatrix}$

\rightarrow

$5 \cdot 3 + 1 \cdot 5 = 20$

\rightarrow

$\begin{bmatrix} 20 & 9 \\ 8 & 1 \end{bmatrix}$

$\begin{bmatrix} 5 & 1 \\ 1 & 1 \end{bmatrix}$

\cdot

$\begin{bmatrix} 3 & 2 \\ 5 & -1 \end{bmatrix}$

\rightarrow

$5 \cdot 2 + 1 \cdot (-1) = 9$

\rightarrow

$\begin{bmatrix} 10 & 9 \\ 8 & 1 \end{bmatrix}$

$\begin{bmatrix} 5 & 1 \\ 1 & 1 \end{bmatrix}$

\cdot

$\begin{bmatrix} 3 & 2 \\ 5 & -1 \end{bmatrix}$

\rightarrow

$1 \cdot 3 + 1 \cdot 5 = 8$

\rightarrow

$\begin{bmatrix} 10 & 9 \\ 8 & 1 \end{bmatrix}$

$\begin{bmatrix} 5 & 1 \\ 1 & 1 \end{bmatrix}$

\cdot

$\begin{bmatrix} 3 & 2 \\ 5 & -1 \end{bmatrix}$

\rightarrow

$1 \cdot 2 + 1 \cdot (-1) = 1$

\rightarrow

$\begin{bmatrix} 10 & 9 \\ 8 & 1 \end{bmatrix}$

In general, we have a single formula:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{bmatrix}$$

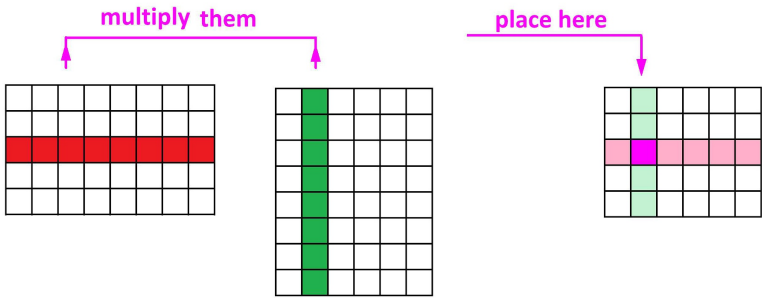
Warning!

As the matrix multiplication is just an abbreviated representation of the composition of linear operators, it is secondary to the substitution of the formulas it comes from.

Now we consider linear operators in arbitrary dimensions:

$$\mathbf{R}^n \rightarrow \mathbf{R}^m \rightarrow \mathbf{R}^k .$$

For their matrices, the length of the column in the first must be equal to the length of the row in the second. That’s m ! Then, we can carry out a dot product:



We do this nk times and produce an $n \times k$ matrix.

Exercise 3.2.6

Multiply:

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \cdot \begin{bmatrix} 1 & -1 \\ -1 & 1 \\ 1 & -1 \end{bmatrix} .$$

Example 3.2.7: spreadsheet

One can utilize a spreadsheet and other software to multiplication for matrices of any dimensions. In order to make this work, the second matrix B has to be “transposed” (bottom):

		A		•	B		=	C		
		1	2	3		1	0		22	8
		2	-1	4	•	3	1	=	19	7
		2	3	5		5	2		36	13
		1	0	2					11	4
				B ^T	=	1	3	5		
						0	1	2		

This is the code for the transpose of X :

=TRANSPOSE(R[-5]C:R[-3]C[1])

This is the code for Y :

=SUMPRODUCT(RC2:RC4,R8C[-3]:R8C[-1])

Finding the inverse of a matrix, especially of high dimension, is a challenging problem. There is a simple formula for the 2×2 matrices:

Theorem 3.2.8: Inverse of 2×2 Matrix

The inverse of an invertible matrix A is computed as follows:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{\det A} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Exercise 3.2.9

Prove the theorem.

Exercise 3.2.10

Use the theorem to solve the following problem: A tourist group took a train trip at \$3 per child and \$3.20 per adult for a total of \$118.40. They took the train back at \$3.50 per child and \$3.60 per adult for a total of \$135.20. How many children, and how many adults?

Exercise 3.2.11

Use the theorem to solve the following problem: A tourist group with 10 children and 20 adults took a train trip for a total of \$110. Another tourist group with 15 children and 25 adults took a train trip for a total of \$145. What were the ticket prices?

To summarize, we go back to the general setup of linear operators applied consecutively:

$$\mathbf{R}^n \xrightarrow{F} \mathbf{R}^m \xrightarrow{G} \mathbf{R}^k$$

We have implicitly used the following fact:

Theorem 3.2.12: Composition of Linear Operators

The composition of two linear operators is a linear operator.

Exercise 3.2.13

Prove the theorem.

We have also implicitly used the the following important result:

Theorem 3.2.14: Matrix of Composition

The product of two matrices that represent two linear operators is the matrix of their composition.

3.3. How complex numbers emerge

The equation

$$x^2 + 1 = 0$$

has no solutions. Indeed, we observe the following:

$$x^2 \geq 0 \implies x^2 + 1 > 0 \implies x^2 + 1 \neq 0.$$

If we try to solve it the usual way, we get these:

$$x = \sqrt{-1} \text{ and } x = -\sqrt{-1}.$$

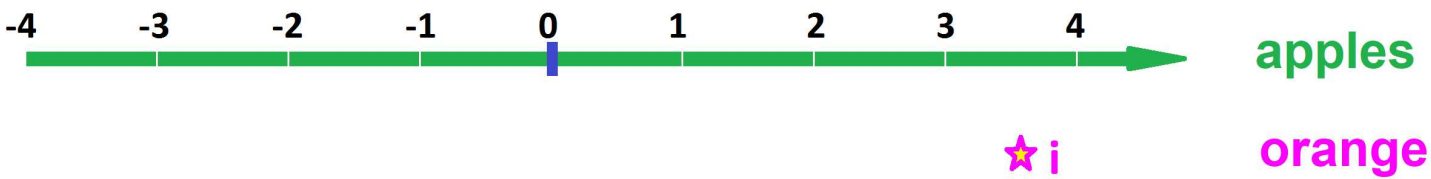
There are no such *real numbers*.

However, let’s ignore this fact for a moment. Let’s substitute what we have back into the equation and – blindly – follow the rules of algebra. We “confirm” that this “number” is a “solution”:

$$x^2 + 1 = (\sqrt{-1})^2 + 1 = (-1) + 1 = 0.$$

We call this entity the *imaginary unit*, denoted by i .

We just add this “number” to the set of numbers we do algebra with:



And see what happens...

Making i a part of algebra will only require this three-part convention:

- 1. i is not a real number (and, in particular, $i \neq 0$), but
- 2. i can participate in the (four) algebraic operations with real numbers by following the same rules; also
- 3. $i^2 = -1$.

What algebraic rules are those? A few very basic ones:

$$x + y = y + x, \ x \cdot y = y \cdot x, \ x(y + z) = xy + xz, \text{ etc.}$$

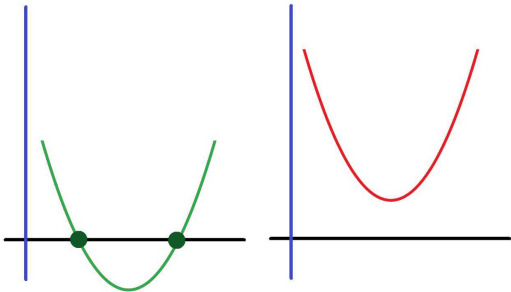
We allow one or several of these parameters to be i . For example, we have:

$$i + y = y + i, \ i \cdot y = y \cdot i, \ i(y + z) = iy + iz, \text{ etc.}$$

What makes this extra effort worthwhile is a new look at quadratic polynomials. For example, this is how we may factor one:

$$x^2 - 1 = (x - 1)(x + 1).$$

Then $x = 1$ and $x = -1$ are the x -intercepts of the polynomial:



But some polynomials, called *irreducible*, cannot be factored; there are no a, b such that:

$$x^2 + 1 = (x - a)(x - b).$$

There are no *real* a, b , that is! Using our rules, we discover:

$$(x - i)(x + i) = x^2 - ix + ix - i^2 = x^2 + 1.$$

Of course, the number i is *not* an x -intercept of $f(x) = x^2 + 1$ as the x -axis (“the real line”) consists of only (and all) real numbers.

So, multiples of i appear immediately as we start doing algebra with it.

Definition 3.3.1: imaginary numbers

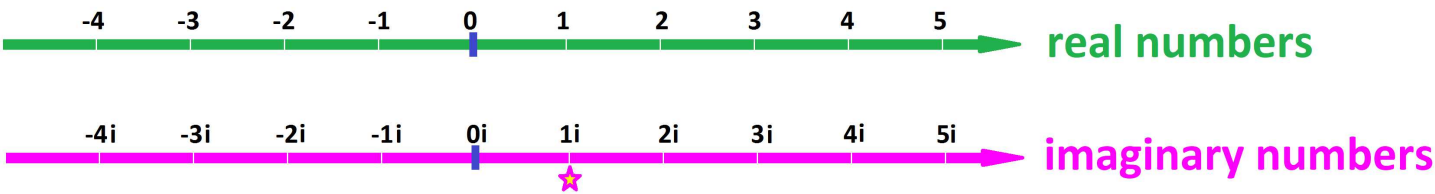
The real multiples of the imaginary unit, i.e.,
$$z = ri, \text{ } r \text{ real,}$$
are called *imaginary numbers*.

We have created a whole class of *non-real* numbers! Of course, ri , where r is real, can’t be real:

$$(ri)^2 = r^2i^2 = -r^2 < 0.$$

The only exception is $0i = 0$; it’s real!

There are as many of them as the real numbers:



Example 3.3.2: quadratic equations

The imaginary numbers may also come from solving the simplest quadratic equations. For example, the equation
$$x^2 + 4 = 0$$
gives us via our substitution:
$$x = \pm\sqrt{-4} = \pm\sqrt{4(-1)} = \pm\sqrt{4}\sqrt{-1} = \pm 2i.$$
Indeed, if we substitute $x = 2i$ into the equation, we have:
$$(2i)^2 + 4 = (2)^2(i)^2 + 4 = 4(-1) + 4 = 0.$$
More general quadratic equations are discussed in the next section.

Imaginary numbers obey the laws of algebra as we know them! If we need to simplify the expression, we try to manipulate it in such a way that real numbers are combined with real while i is pushed aside.

For example, we can just factor i out of all addition and subtraction:

$$5i + 3i = (5 + 3)i = 8i.$$

It looks exactly like middle school algebra:

$$5x + 3x = (5 + 3)x = 8x.$$

After all, x *could* be i . Another similarity is with the algebra of quantities that have units:

$$5 \text{ in.} + 3 \text{ in.} = (5 + 3) \text{ in.} = 8 \text{ in..}$$

So, the nature of the unit doesn’t matter (if we can push it aside). Even simpler:

$$5 \text{ apples} + 3 \text{ apples} = (5 + 3) \text{ apples} = 8 \text{ apples}.$$

It’s “8 apples” not “8”! And so on.

This is how we multiply an imaginary number by a real number:

$$2 \cdot (3i) = (2 \cdot 3)i = 6i .$$

We have a new imaginary number.

How do we multiply two imaginary numbers? It’s different; after all, we don’t usually multiply apples by apples! In contrast to the above, even though multiplication and division follow the same rule as always, we can, when necessary, and often have to, simplify the outcome of our algebra using our *fundamental identity*:

$$i^2 = -1 .$$

For example:

$$(5i) \cdot (3i) = (5 \cdot 3)(i \cdot i) = 15i^2 = 15(-1) = -15 .$$

It’s real!

We also simplify the outcome by using the other *fundamental fact* about the imaginary unit:

$$i \neq 0 .$$

We can divide by i ! For example,

$$\frac{5i}{3i} = \frac{5}{3} \frac{i}{i} = \frac{5}{3} \cdot 1 = \frac{5}{3} .$$

As you can see, doing algebra with imaginary numbers will often bring us back to real numbers. These two classes of numbers cannot be separated from each other!

They aren’t. Let’s take another look at quadratic equations. The equation

$$ax^2 + bx + c = 0, \ a \neq 0 ,$$

is solved with the familiar *Quadratic Formula*:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} .$$

Let’s consider

$$x^2 + 2x + 10 = 0 .$$

Then the roots are supposed to be:

$$\begin{aligned} x &= \frac{-2 \pm \sqrt{2^2 - 4 \cdot 10}}{2} \\ &= \frac{-2 \pm \sqrt{-36}}{2} \\ &= -1 \pm \sqrt{-9} && \text{There is no real solution!} \\ &= -1 \pm \sqrt{9}\sqrt{-1} && \text{But we go on.} \\ &= -1 \pm 3i . \end{aligned}$$

We end up adding real and imaginary numbers!

As there is no way to simplify this, we conclude the following:

- A number $a + bi$, where $a, b \neq 0$ are real, is neither real nor imaginary.

Exercise 3.3.3

Explain why.

This addition is not literal. It’s like “adding” apples to oranges:

$$5 \text{ apples} + 3 \text{ oranges} = \dots$$

It’s not 8 and it’s not 8 fruit because we wouldn’t be able to read this equality backwards. The algebra will, however, be meaningful:

$$(5a + 3o) + (2a + 4o) = (5 + 3)a + (3 + 4)o = 8a + 7o.$$

It is as if we collect *similar terms*, like this:

$$(5 + 3x) + (2 + 4x) = (5 + 2) + (3 + 4)x = 8 + 7x.$$

This idea enables us to do this:

$$(5 + 3i) + (2 + 4i) = (5 + 3) + (3 + 4)i = 8 + 7i.$$

Each of the numbers we are facing contain both real numbers and imaginary parts. This fact makes them “complex”...

Definition 3.3.4: complex number

Any sum of real and imaginary numbers is called a *complex number*. The set of all complex numbers is denoted as follows:

$$\mathbf{C} = \{z = a + bi : a, b \text{ real}\}$$

Warning!

All real numbers are complex ($b = 0$).

Addition and subtraction are easy; we just combine *similar terms* just like in middle school. For example,

$$(1 + 5i) + (3 - i) = 1 + 5i + 3 - i = (1 + 3) + (5i - i) = 4 + 4i.$$

To simplify *multiplication* of complex numbers, we expand and then use $i^2 = -1$, as follows:

$$\begin{aligned}(1 + 5i) \cdot (3 - i) &= 1 \cdot 3 + 5i \cdot 3 + 1 \cdot (-i) + 5i \cdot (-i) \\ &= 3 + 15i - i - 5i^2 \\ &= (3 + 5) + (15i - i) \\ &= 8 + 14i.\end{aligned}$$

It’s a bit trickier with *division*:

$$\begin{aligned}\frac{1 + 5i}{3 - i} &= \frac{1 + 5i}{3 - i} \frac{3 + i}{3 + i} \\ &= \frac{(1 + 5i)(3 + i)}{(3 - i)(3 + i)} \\ &= \frac{-2 + 8i}{3^2 - i^2} \\ &= \frac{-2 + 8i}{3^2 + 1} \\ &= \frac{1}{10}(-2 + 8i) \\ &= -0.2 + 0.8i.\end{aligned}$$

The simplification of the denominator is made possible by the trick of multiplying by $3 + i$. It is the same trick we used in Volume 1 to simplify fractions with roots to compute their limits:

$$\frac{1}{1 - \sqrt{x}} = \frac{1}{1 - \sqrt{x}} \frac{1 + \sqrt{x}}{1 + \sqrt{x}} = \frac{1 + \sqrt{x}}{1 - x}.$$

Definition 3.3.5: complex conjugate

The *complex conjugate* of $z = a + bi$ is defined and denoted as follows:

$$\bar{z} = \overline{a + bi} = a - bi.$$

The following is crucial.

Theorem 3.3.6: Algebra of Complex Numbers

The rules of the algebra of complex numbers are identical to those of real numbers:

- **Commutativity of addition:** $z + u = u + z$
- **Associativity of addition:** $(z + u) + v = z + (u + v)$
- **Commutativity of multiplication:** $z \cdot u = u \cdot z$
- **Associativity of multiplication:** $(z \cdot u) \cdot v = z \cdot (u \cdot v)$
- **Distributivity:** $z \cdot (u + v) = z \cdot u + z \cdot v$

This is the *complex number system*; it follows the rules of the real number system but also contains it. This theorem will allow us to build calculus for complex functions that is almost identical to calculus for real functions but also contains it.

Definition 3.3.7: standard form of complex number

Every complex number x has the *standard representation*:

$$z = a + bi,$$

where a and b are two real numbers. The two components are named as follows:

- a is the *real part* of z , with notation:

$$a = \operatorname{Re}(z);$$

- bi is the *imaginary part* of z , with notation:

$$b = \operatorname{Im}(z).$$

Then, the purpose of the computations above was to find the standard form of a complex number that comes from algebraic operations with other complex numbers. They were literally simplifications.

The definition makes sense because of the following result:

Theorem 3.3.8: Standard Form of Complex Number

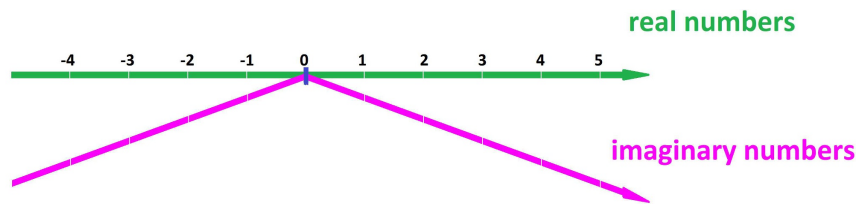
Two complex numbers are equal if and only if both their real and imaginary parts are equal.

So, we have:

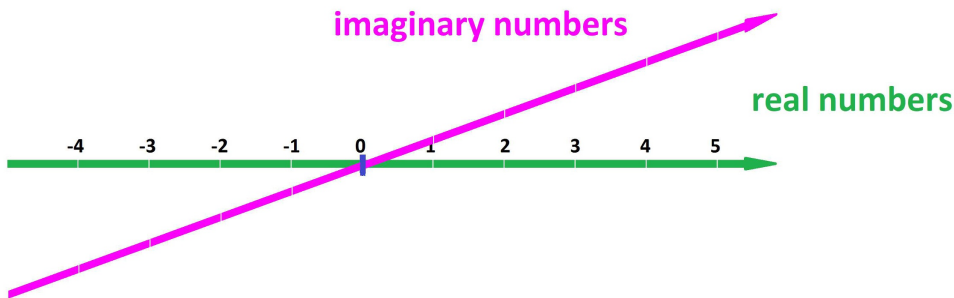
$$z = \operatorname{Re}(z) + \operatorname{Im}(z)i.$$

In order to see the *geometric representation of complex numbers*, we need to combine the real number line and the imaginary number line. How? We realize that they have nothing in common... except $0 = 0i$

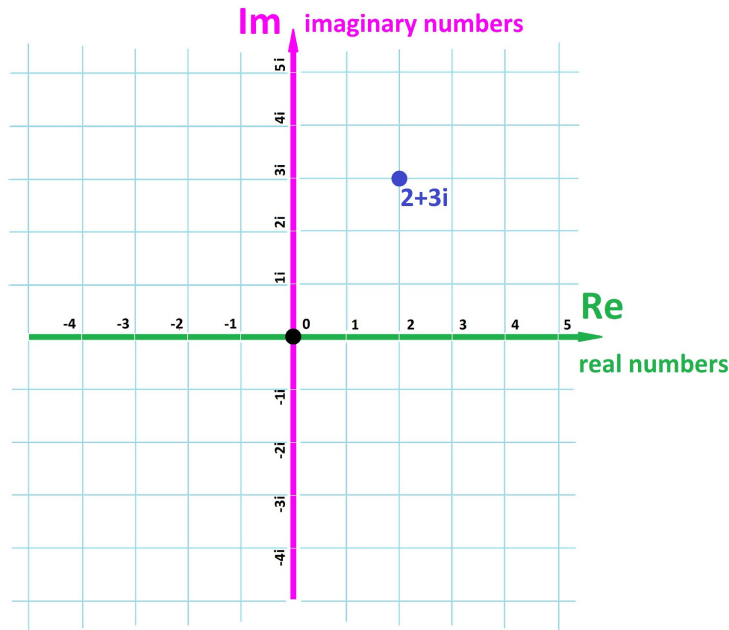
belongs to both:



We can try to combine them like that, or like this:



Or we can try to combine them in the same manner we built the xy -plane:



This representation helps us understand the main idea:

- Complex numbers are linear combinations of the real unit, 1, and the imaginary unit, i .

If $z = a + bi$, then a and b are thought of as the components of vector z in the plane. We have a one-to-one correspondence:

$$\mathbb{C} \longleftrightarrow \mathbb{R}^2,$$

given by

$$a + bi \longleftrightarrow \langle a, b \rangle .$$

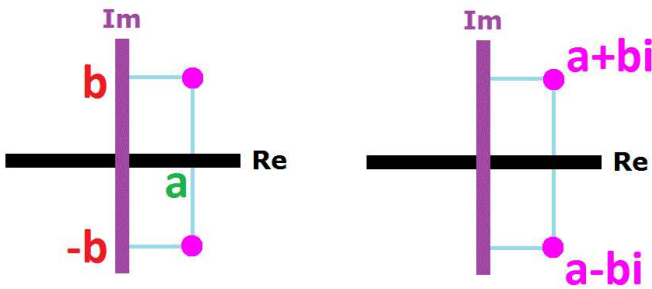
Then the x -axis of this plane consists of the real numbers and the y -axis of the imaginary numbers. It is called the *complex plane*.

Warning!

This is just a visualization.

Then the complex conjugate of z is the complex number with the same real part as z and the imaginary part with the opposite sign:

$$\operatorname{Re}(\bar{z}) = \operatorname{Re}(z) \quad \text{and} \quad \operatorname{Im}(\bar{z}) = -\operatorname{Im}(z).$$



Warning!

All numbers we have encountered so far are real non-complex, and so are all quantities one can encounter in day-to-day life or science: time, location, length, area, volume, mass, temperature, money, etc.

3.4. Classification of quadratic polynomials

The general quadratic equation with real coefficients,

$$ax^2 + bx + c = 0, \quad a \neq 0,$$

can be simplified. Let's divide by a and study the resulting quadratic polynomial:

$$f(x) = x^2 + px + q,$$

where $p = b/a$ and $q = c/a$. The *Quadratic Formula* then provides the x -intercepts of this function:

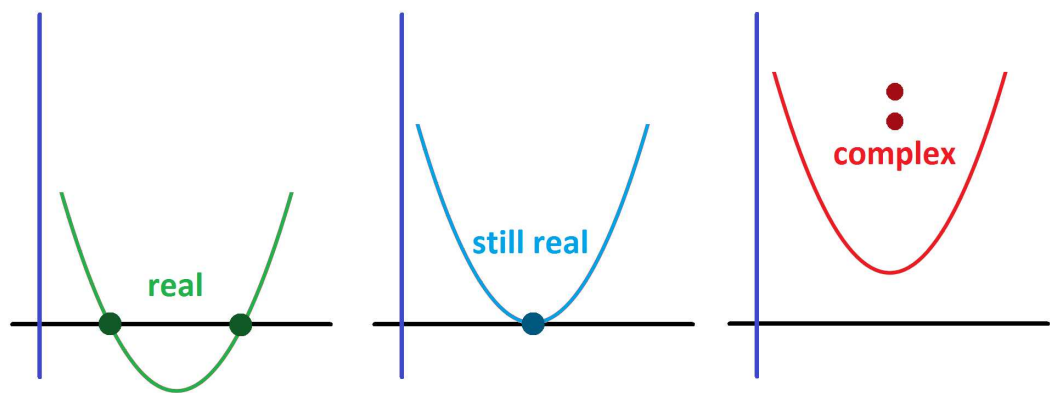
$$x = -\frac{p}{2} \pm \frac{\sqrt{p^2 - 4q}}{2}.$$

Of course, the x -intercepts are the real solutions of this equation and that is why the result only makes sense when the *discriminant* of the quadratic polynomial,

$$D = p^2 - 4q,$$

is non-negative.

Now, increasing the value of q makes the graph of $y = f(x)$ shift upward and, eventually, pass the x -axis entirely. We can observe how its two x -intercepts start to get closer to each other, then merge, and finally disappear:



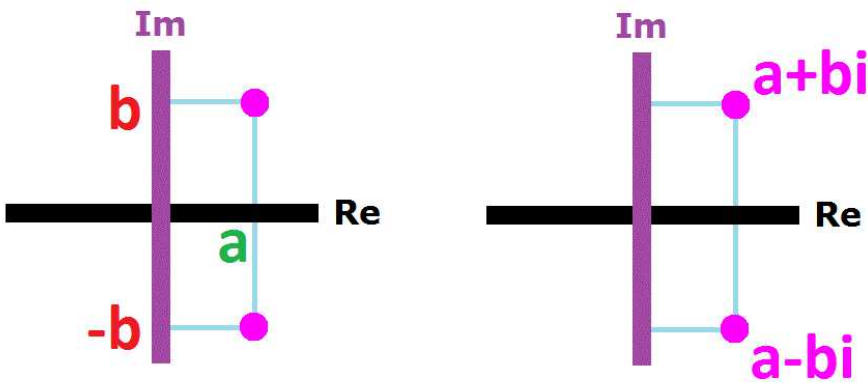
This process is explained by what is happening, with the growth of q , to the roots given by the *Quadratic Formula*:

$$x_{1,2} = -\frac{p}{2} \pm \frac{\sqrt{D}}{2}.$$

- Starting with a positive value, D decreases and $\frac{\sqrt{D}}{2}$ decreases; then
- $D = 0$ and $\frac{\sqrt{D}}{2} = 0$; then
- D becomes negative and $\frac{\sqrt{D}}{2}$ becomes imaginary (but $\frac{\sqrt{-D}}{2}$ is real). The roots are, respectively:

discriminant	root #1	root #2
$D > 0$	$x_1 = -\frac{p}{2} - \frac{\sqrt{D}}{2}$	$x_2 = -\frac{p}{2} + \frac{\sqrt{D}}{2}$
$D = 0$	$x_1 = -\frac{p}{2}$	$x_2 = -\frac{p}{2}$
$D < 0$	$x_1 = -\frac{p}{2} - \frac{\sqrt{-D}}{2}i$	$x_2 = -\frac{p}{2} + \frac{\sqrt{-D}}{2}i$

Observe that the real roots ($D > 0$) are unrelated while the complex ones ($D < 0$) are linked so much that knowing one tells us what the one is: just flip the sign; they are *conjugate* of each other:



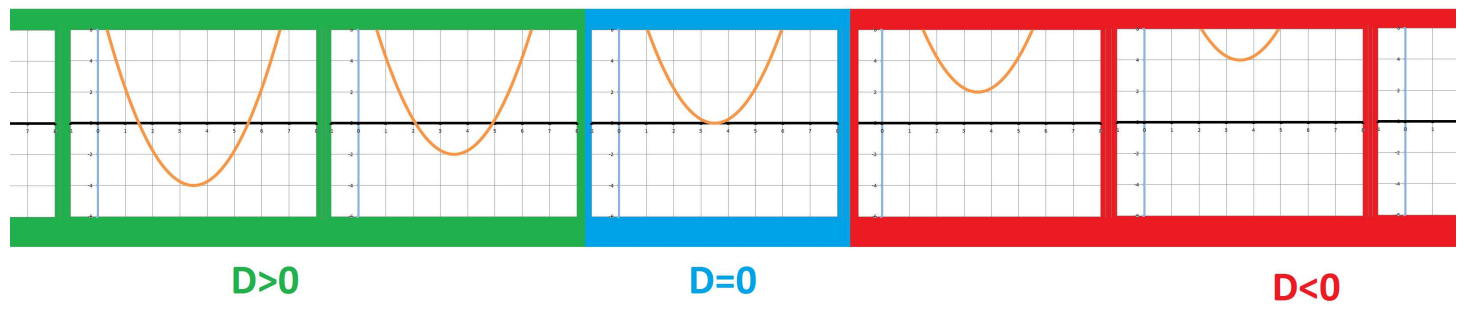
They always come in pairs!

As a summary, we have the following classification the roots of quadratic polynomials in terms of *the sign of the discriminant*.

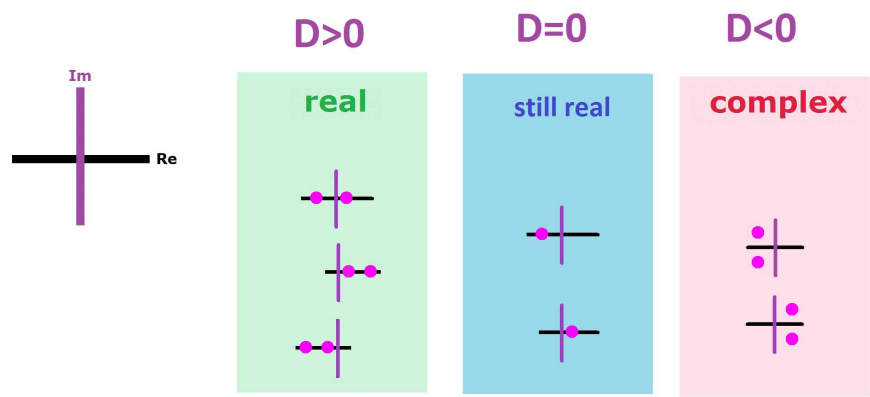
Theorem 3.4.1: Classification of Roots I

The two roots of a quadratic polynomial with real coefficients are:

- distinct real when its discriminant D is positive;
- equal real when its discriminant D is zero;
- complex conjugate of each other when its discriminant D is negative.



To understand ODEs, we will need a more precise way to classify the polynomials: according to *the signs of the real parts of their roots*. The signs will determine increasing and decreasing behavior of certain solutions. Once again, these are the possibilities:



Theorem 3.4.2: Classification of Roots II

Suppose x_1 and x_2 are the two roots of a quadratic polynomial $f(x) = x^2 + px + q$ with real coefficients. Then the signs of the real parts $\operatorname{Re}(x_1)$ and $\operatorname{Re}(x_2)$ of x_1 and x_2 are:

- same when $p^2 > 4q$ and $q \geq 0$;
- opposite when $p^2 > 4q$ and $q < 0$;
- same and opposite of that of p when $p^2 \leq 4q$.

Proof.

The condition $p^2 \leq 4q$ is equivalent to $D \leq 0$. We can see in the table above that, in that case, we have $\operatorname{Re}(x_1) = \operatorname{Re}(x_2) = -\frac{p}{2}$. We are left with the case $D > 0$ and real roots. The case of equal signs of x_1 and x_2 is separated from the case of opposite signs of x_1 and x_2 by the case when both are equal to zero: $x_1 = x_2 = 0$. We solve:

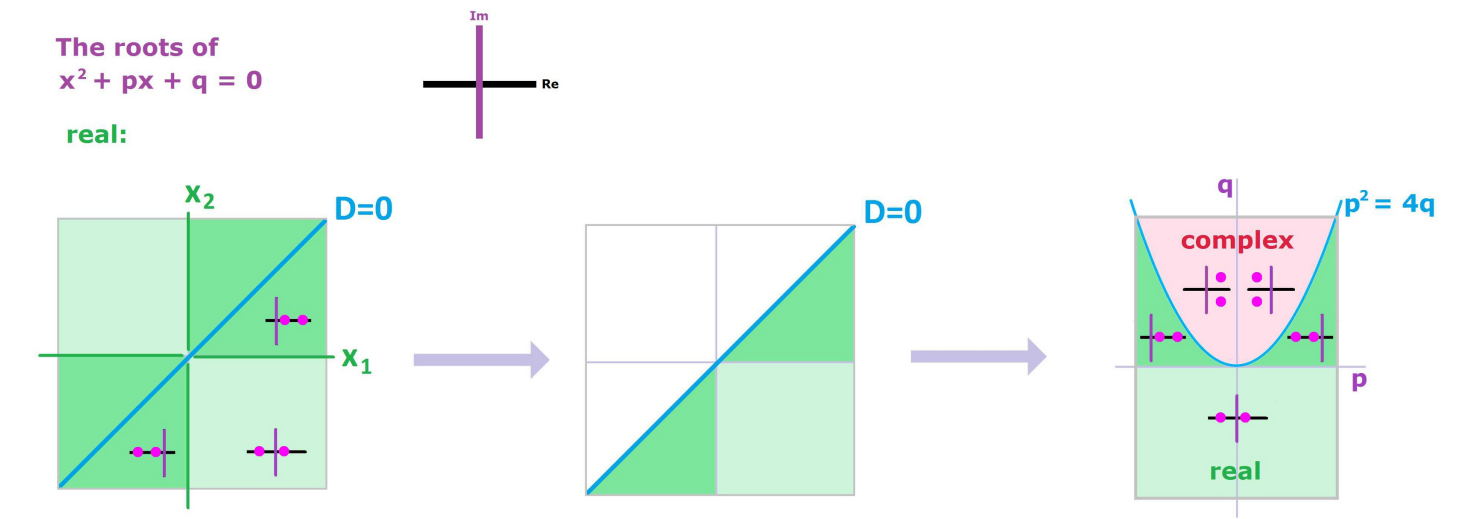
$$-\frac{p}{2} - \frac{\sqrt{D}}{2} = 0 \implies p = -\sqrt{p^2 - 4q} \implies p^2 = p^2 - 4q \implies q = 0.$$

Exercise 3.4.3

Finish the proof.

Let’s visualize our conclusion. We would like to show the main scenarios of what kinds of roots the polynomial might have depending on the values of its two coefficients, p and q .

First, how do we visualize *pairs of numbers*? As points on a coordinate plane of course... but only when they are *real*. Suppose for now that they are. We start with a plane, the x_1x_2 -plane to be exact, as a representation of all possible pairs of real roots (left). Then the diagonal of this plane represents the case of *equal* (and still real) roots, $x_1 = x_2$, i.e., $D = 0$. Since the order of the roots doesn’t matter – (x_1, x_2) is as good as (x_2, x_1) – we need only half of the plane. We fold the plane along the diagonal (middle).



The diagonal – represented by the equation $D = 0$ – exposed this way can now serve its purpose of separating the case of real and *complex* roots. Now, let’s go to the pq -plane. Here, the parabola $p^2 = 4q$ also represents the equation $D = 0$. Let’s bring them together! We take our half-plane and bend its diagonal edge into the parabola $p^2 = 4q$ (right).

Classifying polynomials this way allows one to classify matrices and understand what each of them does as a transformation of the plane, which in turn will help us understand systems of ODEs.

3.5. The complex plane \mathbb{C} is the Euclidean space \mathbb{R}^2

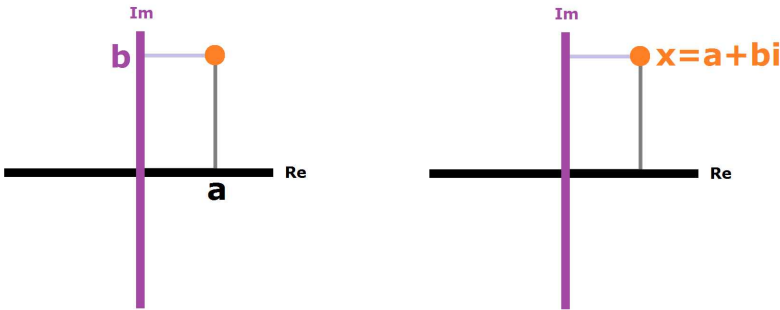
If we call complex number *numbers*, they must be subject to some *algebraic operations*.

We will initially look at them through the lens of *vector algebra* of the plane \mathbb{R}^2 .

A complex number z has the *standard representation*:

$$z = a + bi,$$

where a and b are two real numbers. These two can be seen in the *geometric representation* of complex numbers:



Therefore, a and b are thought of as the coordinates of z as a *point* on the plane. But any complex number is not only a point on the complex plane but also a *vector*. We have a correspondence:

$$\mathbb{C} \longleftrightarrow \mathbb{R}^2,$$

given by

$a + bi \longleftrightarrow \langle a, b \rangle$

There is more to this than just a match; the algebra of vectors in \mathbb{R}^2 applies!

Warning!

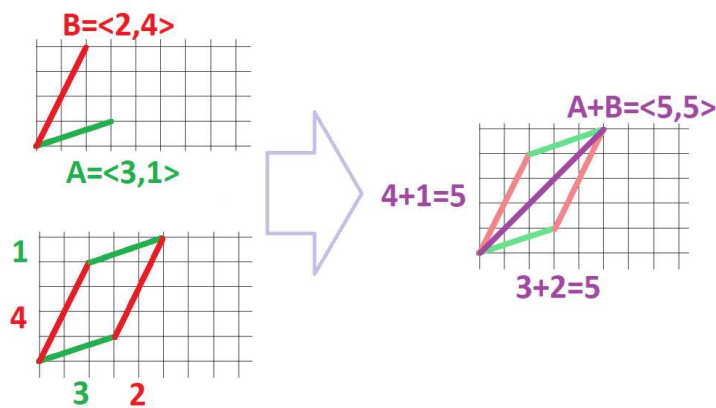
In spite of this fundamental correspondence, we will continue to think of complex numbers as *numbers* (and use the lower case letters).

Let’s see how this algebra of numbers works in parallel with the algebra of 2-vectors.

First, the addition of complex numbers is done *componentwise*:

$$\begin{aligned} (a + bi) + (c + di) &= (a + c) + (b + d)i \\ \langle a, b \rangle + \langle c, d \rangle &= \langle a + c, b + d \rangle \end{aligned}$$

It corresponds to addition of vectors:



Second, we can easily multiply complex numbers by real ones:

$$\begin{aligned} (a + bi) \cdot c &= (ac) + (bc)i \\ \langle a, b \rangle \cdot c &= \langle ac, bc \rangle \end{aligned}$$

It corresponds to scalar multiplication of vectors.

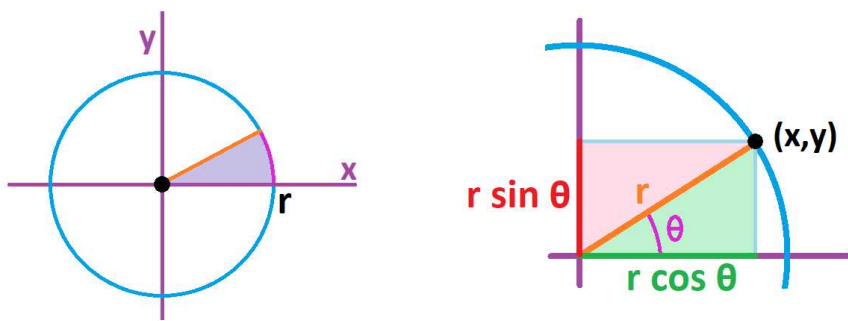
Warning!

Vector algebra of \mathbb{R}^2 is complex algebra, but not vice versa. Complex *multiplication* is what makes it different.

Example 3.5.1: circle

We can easily represent circles on the complex plane:

$$z = r \cos \theta + r \sin \theta \cdot i.$$



Our study of calculus of complex numbers starts with the study of the *topology* of the complex plane. This topology is the same as that of the *Euclidean plane* \mathbb{R}^2 !

Just as before, every function $z = f(t)$ with an appropriate domain creates a sequence:

$$z_k = f(k) .$$

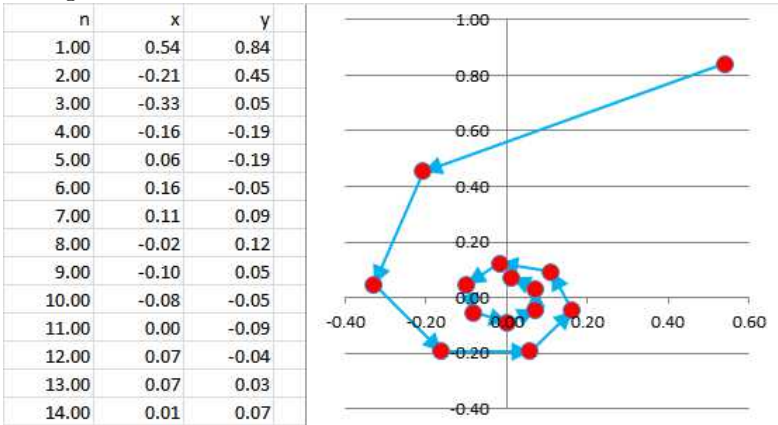
A function with complex values defined on a ray in the set of integers, $\{p, p + 1, \dots\}$, is called an *infinite sequence*, or simply *sequence*.

Example 3.5.2: spiral

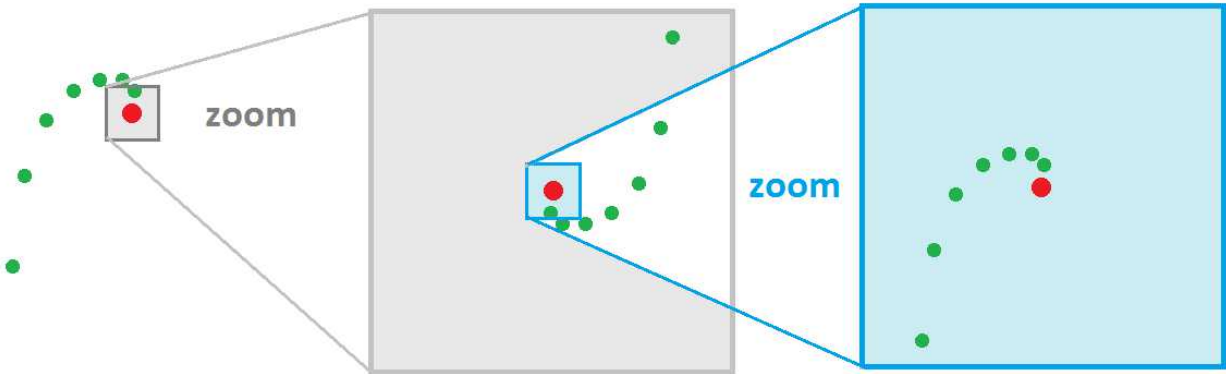
A good example is that of the sequence made of the reciprocals:

$$z_k = \frac{\cos k}{k} + \frac{\sin k}{k}i .$$

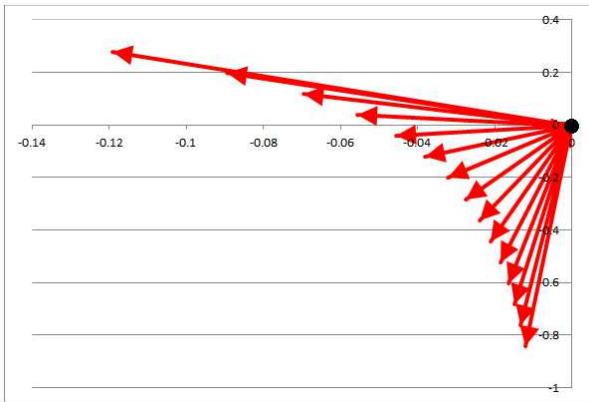
It *tends to 0* while spiraling around it.



The starting point of calculus of complex numbers is the following. The convergence of a *sequence of complex numbers* is the convergence of its real and imaginary parts or, which is equivalent, the convergence of points (or vectors) on the complex plane seen as any plane: The distance from the k th point to the limit is getting smaller and smaller.



We use the definition of convergence for vectors on the plane by simply replacing vectors with complex numbers and “magnitude” with “modulus”.



Definition 3.5.3: convergent sequence

Suppose $\{z_k : k = 1, 2, 3, \dots\}$ is a sequence of complex numbers, i.e., points in \mathbf{C} . We say that the sequence *converges* to another complex number z , i.e., a point in \mathbf{C} , called the *limit* of the sequence, if:

$$||z_k - z|| \rightarrow 0 \text{ as } k \rightarrow \infty,$$

denoted as follows:

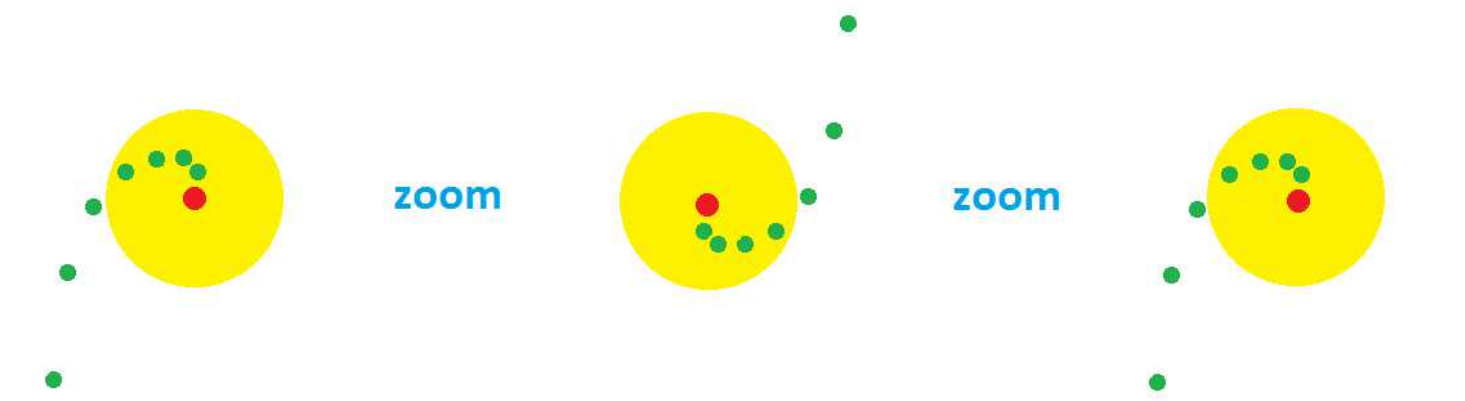
$$z_k \rightarrow z \text{ as } k \rightarrow \infty,$$

or

$$z = \lim_{k \rightarrow \infty} z_k.$$

If a sequence has a limit, we call the sequence *convergent* and say that it *converges*; otherwise it is *divergent* and we say it *diverges*.

In other words, the points start to accumulate in smaller and smaller circles around z . A way to visualize a trend in a convergent sequence is to enclose the tail of the sequence in a *disk*:



Theorem 3.5.4: Uniqueness of Limit

A sequence can have only one limit (finite or infinite); i.e., if a and b are limits of the same sequence, then $a = b$.

Definition 3.5.5: sequence tends to infinity

We say that a sequence z_k *tends to infinity* if the following condition holds: For each real number R , there exists such a natural number N that, for every natural number $k > N$, we have

$$||z_k|| > R.$$

We use the following notation:

$$z_k \rightarrow \infty \text{ as } k \rightarrow \infty.$$

The following is another analog of a familiar theorem about the topology of the plane.

Theorem 3.5.6: Componentwise Convergence of Sequences

A sequence of complex numbers z_k in \mathbf{C} converges to a complex number z if and only if both the real and the imaginary parts of z_k converge to the real and the imaginary parts of z respectively; i.e.,

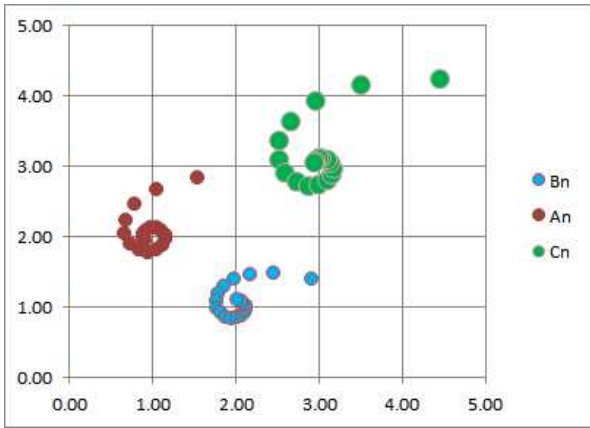
$$z_k \rightarrow z \iff \operatorname{Re}(z_k) \rightarrow \operatorname{Re}(z) \text{ and } \operatorname{Im}(z_k) \rightarrow \operatorname{Im}(z).$$

The algebraic properties of limits of sequences of complex numbers will also look familiar:

Theorem 3.5.7: Sum Rule for Complex Sequences

If sequences z_k, u_k converge, then so does $z_k + u_k$, and we have:

$$\lim_{k \rightarrow \infty} (z_k + u_k) = \lim_{k \rightarrow \infty} z_k + \lim_{k \rightarrow \infty} u_k .$$



Theorem 3.5.8: Constant Multiple Rule for Complex Sequences

If sequence z_k converges, then so does cz_k for any complex number c , and we have:

$$\lim_{k \rightarrow \infty} c z_k = c \cdot \lim_{k \rightarrow \infty} z_k .$$

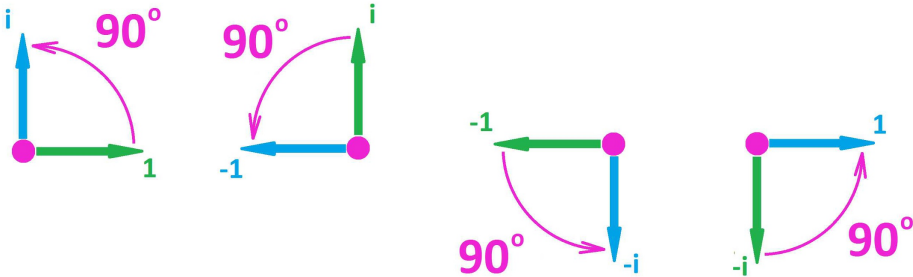
Wouldn't calculus of complex numbers be just a copy of calculus on the plane? No, not with the possibility of *multiplication* taken into account.

3.6. Multiplication of complex numbers: \mathbf{C} isn't just \mathbf{R}^2

So, the vector algebra of \mathbf{R}^2 is included in the complex algebra of \mathbf{C} . There is more to the latter. Just like in \mathbf{R}^2 , multiplication by a *real* number r will stretch/shrink all vectors and, therefore, the complex plane \mathbf{C} . However, multiplication by a *complex* number c will also rotate each vector.

Example 3.6.1: multiplication by i

Let's start with 1 and multiply it by i several times. Multiplication by i rotates the number by 90 degrees: 1 becomes i , while i becomes -1 , etc.:



$1 \cdot i = i$	rotation from 0 degrees to 90
$i \cdot i = i^2 = -1$	rotation from 90 degrees to 180
$-1 \cdot i = -i$	rotation from 180 degrees to 270
$-i \cdot i = -i^2 = 1$	rotation from 270 degrees to 360
	and so on.

Example 3.6.2: complex multiplication

A more complex example:

u

$= 1 + 2i$

v

$= 2 + i$

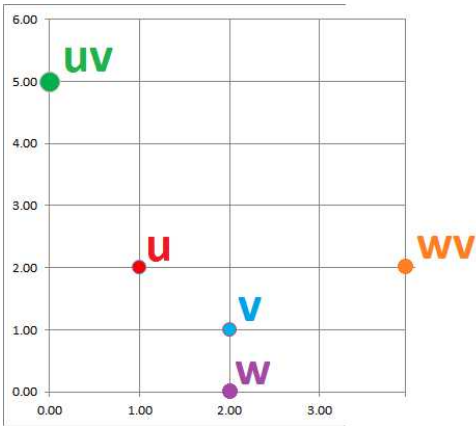
uv

$= 2 + 4i + i + 2i^2$

$= (2 - 2) + (4 + 1)i$

$= 0 + 5i$

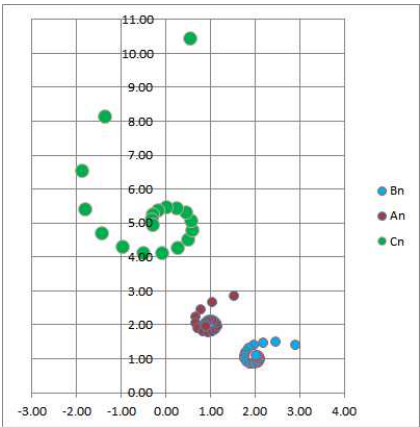
The rotation of v is visible:



In contrast, we can see the result of multiplying v by $w = 2$: no rotation.

So, the imaginary part of c is responsible for rotation.

How does multiplication affect topology?



Theorem 3.6.3: Product Rule for Complex Sequences

If sequences z_k, u_k converge, then so does $z_k \cdot u_k$, and

$$\lim_{k \rightarrow \infty} (z_k \cdot u_k) = \lim_{k \rightarrow \infty} z_k \cdot \lim_{k \rightarrow \infty} u_k .$$

Proof.

Suppose

$$z_k = a_k + b_k i \rightarrow a + bi \quad \text{and} \quad u_k = p_k + q_k i \rightarrow p + qi .$$

Then, according to the *Componentwise Convergence Theorem* above, we have:

$$a_k \rightarrow a, \; b_k \rightarrow b \; \text{ and } \; p_k \rightarrow p, \; q_k \rightarrow q .$$

Then, by the *Product Rule for numerical sequences*, we have:

$$a_k p_k \rightarrow ap, \; a_k q_k \rightarrow aq, \; b_k p_k \rightarrow bp, \; b_k q_k \rightarrow bq .$$

Then, as we know,

$$z_k \cdot u_k = (a_k p_k - b_k q_k) + (a_k q_k + b_k p_k)i \rightarrow (ap - bq) + (aq + bq)i = (a + bi)(p + qi) ,$$

by the *Sum Rule for numerical sequences*.

Theorem 3.6.4: Quotient Rule for Complex Sequences

If sequences $z_k, \; u_k$ converge (with $u_k \neq 0$), then so does z_k/u_k , and

$$\lim_{k \rightarrow \infty} \frac{z_k}{u_k} = \frac{\lim_{k \rightarrow \infty} z_k}{\lim_{k \rightarrow \infty} u_k} ,$$

provided

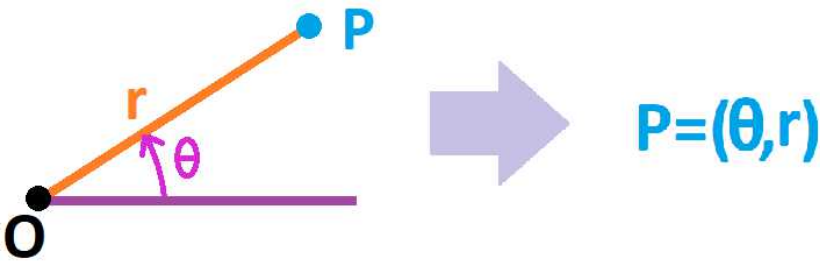
$$\lim_{k \rightarrow \infty} u_k \neq 0 .$$

Just like real numbers!

Exercise 3.6.5

Prove the last theorem.

In addition to the standard, Cartesian, representation, a complex number $x = a + bi$ can be defined in terms of the *polar coordinates*.



We just append our correspondence with a new one:

$$a + bi \longleftrightarrow (a, b) \longleftrightarrow (\theta, r)$$

The two quantities θ and r become the following:

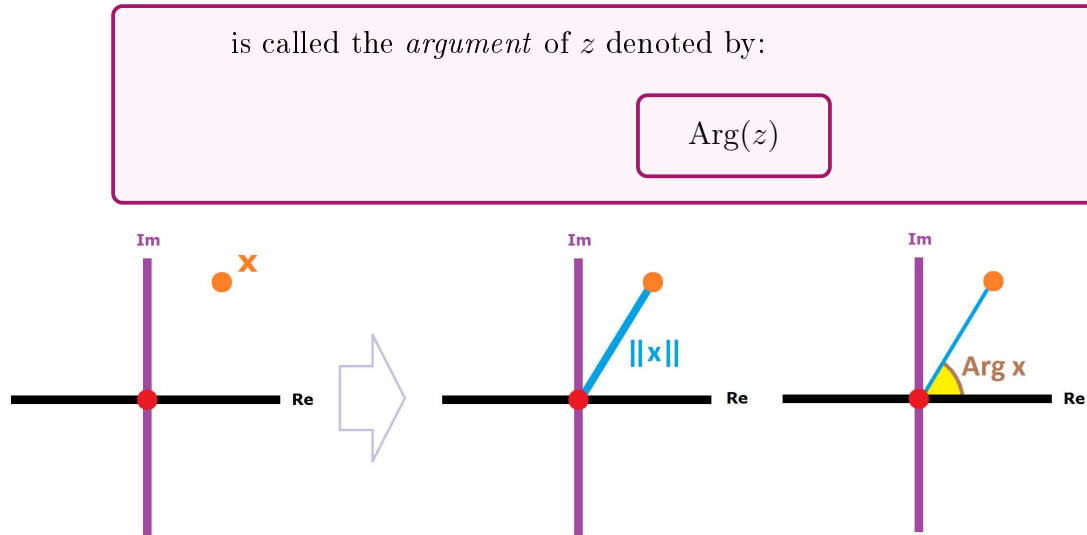
Definition 3.6.6: modulus and argument

Suppose z is a complex number.

1. The distance from the location of z on the complex plane to the origin O is called the *modulus* of z denoted by:

$$||z||$$

2. The angle of the line through the point z from the origin O with the x -axis



A simple examination tells us how to transition between the two coordinate systems:

Theorem 3.6.7: Conversion of Complex Numbers

Suppose $x = a + bi$ is a complex number. Then, we have:

- The modulus of z is found by:

$$||z|| = \sqrt{a^2 + b^2}$$

- The argument of z is found by:

$$\text{Arg}(z) = \arctan \frac{b}{a}$$

Any two real numbers $r \geq 0$ and $0 \leq \theta < 2\pi$ can serve as those. It is called the *geometric representation* of the complex number:

$$z = r [\cos \theta + i \sin \theta]$$

The *algebra* takes a new form too. We don't need the new representation to compute addition and multiplication by real numbers, but we *need* it for multiplication.

What is the product of two complex numbers:

$$z_1 = r_1 [\cos \varphi_1 + i \sin \varphi_1] \quad \text{and} \quad z_2 = r_2 [\cos \varphi_2 + i \sin \varphi_2] ?$$

Consider:

$$\begin{aligned} z_1 z_2 &= r_1 [\cos \varphi_1 + i \sin \varphi_1] \cdot r_2 [\cos \varphi_2 + i \sin \varphi_2] \\ &= r_1 r_2 [\cos \varphi_1 + i \sin \varphi_1] \cdot [\cos \varphi_2 + i \sin \varphi_2] \\ &= r_1 r_2 (\cos \varphi_1 \cos \varphi_2 + i \sin \varphi_1 \cos \varphi_2 + \cos \varphi_1 i \sin \varphi_2 + i^2 \sin \varphi_1 \sin \varphi_2) . \end{aligned}$$

We utilize the following trigonometric identities ([Volume 1](#)):

$$\cos a \cos b - \sin a \sin b = \cos(a + b) \quad \text{and} \quad \cos a \sin b + \sin a \cos b = \sin(a + b) .$$

Then,

$$z_1 z_2 = r_1 r_2 [\cos(\varphi_1 + \varphi_2) + i \sin(\varphi_1 + \varphi_2)] .$$

Example 3.6.8: geometric representation of multiplication

We can see the above computation on the complex plane:

We have proven the following:

Theorem 3.6.9: Multiplication of Complex Numbers

When two complex numbers are multiplied, their moduli are multiplied and the arguments are added.

In other words, we have:

$$\begin{aligned} & r_1 [\cos \varphi_1 + i \sin \varphi_1] \cdot r_2 [\cos \varphi_2 + i \sin \varphi_2] \\ &= r_1 r_2 [\cos(\varphi_1 + \varphi_2) + i \sin(\varphi_1 + \varphi_2)] \end{aligned}$$

Exercise 3.6.10

Suppose we have a convergent sequence of complex numbers. Consider the sequence of the moduli and the sequence of the arguments of the terms of the sequence and prove that they converge.

Exercise 3.6.11

(a) Represent the following complex number in the standard form: $(2 + 3i)(-1 + 2i)$. Indicate the real and imaginary parts. (b) Find its modulus and argument.

Exercise 3.6.12

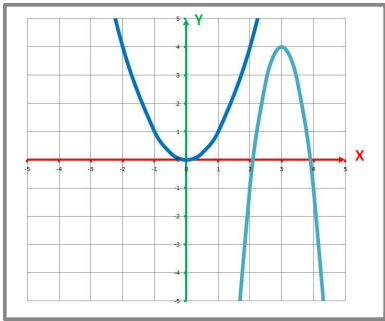
Simplify $(1 + i)^2$.

Exercise 3.6.13

(a) Find the roots of the polynomial $x^2 + 2x + 2$. (b) Find its x -intercepts. (c) Find its factors.

Exercise 3.6.14

What can you say about the imaginary parts of the roots of these quadratic polynomials?



3.7. Complex functions

A complex function is simply a function with both input and output complex numbers:

$$F : \mathbb{C} \rightarrow \mathbb{C} .$$

How do we visualize these functions? The graph of a function F lies in the 4-dimensional space and isn't of much help!

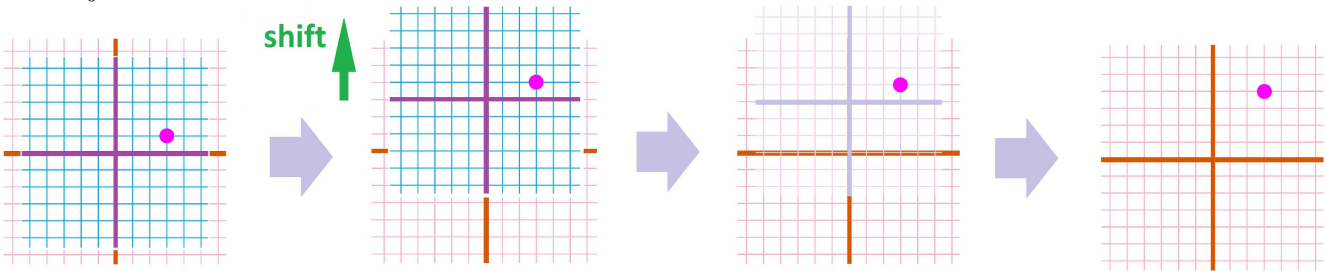
To begin with, we can recast some of the transformations of the plane presented in this chapter as complex functions.

Example 3.7.1: complex addition

The *shift by vector* $V = \langle a, b \rangle$ becomes addition of a fixed complex number:

$$F(x, y) = (x + a, y + b) \text{ , re-written } F(z) = z + z_0 ,$$

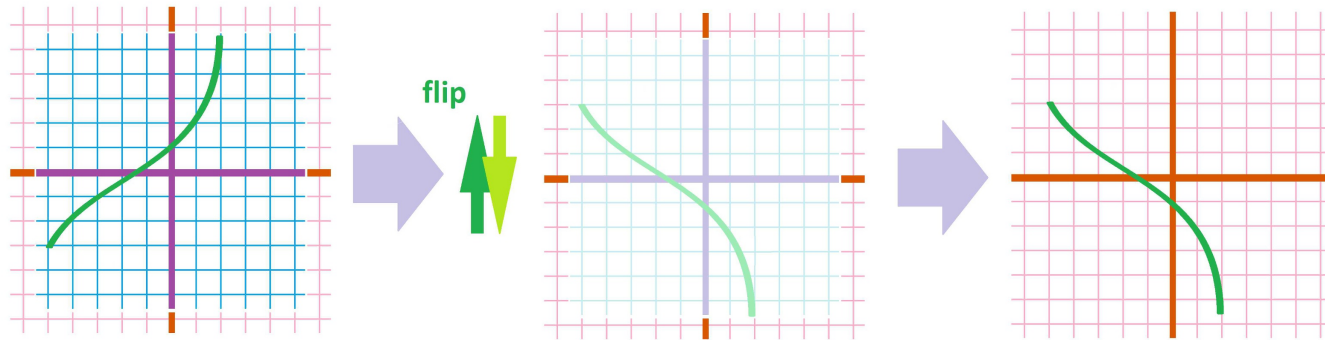
where $z_0 = a + bi$.



Example 3.7.2: complex conjugation

The *vertical flip* becomes conjugation:

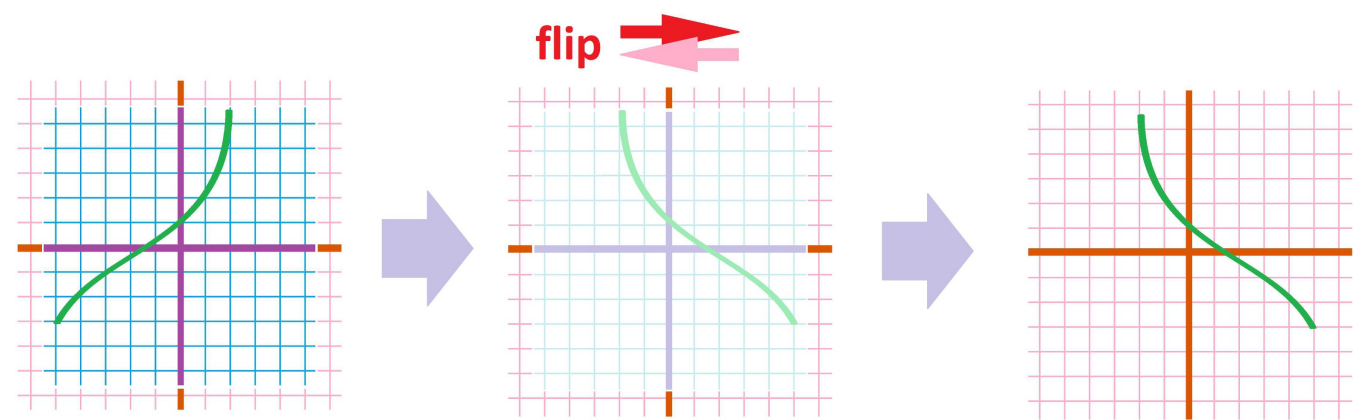
$$F(x, y) = (x, -y) \text{ , re-written } F(z) = \bar{z} .$$



Exercise 3.7.3

Find a complex formula for the *horizontal flip*:

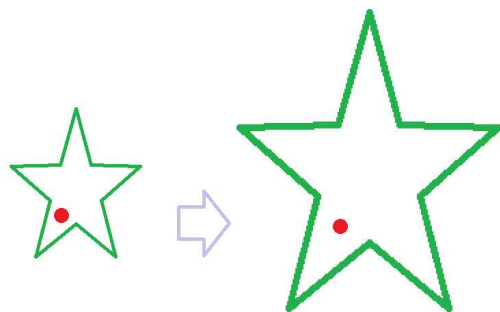
$$F(x,y) = (-x,y) .$$



Example 3.7.4: complex multiplication

The *uniform stretch* becomes multiplication by a *real* number:

$$F(x,y) = (kx,ky) , \text{ re-written } F(z) = kz .$$



Of course, the *flip* about the origin is just multiplication by -1 .

Exercise 3.7.5

Find complex formulas for the *vertical stretch*:

$$F(x,y) = (x,ky)$$

and the horizontal stretch:

$$F(x,y) = (kx,y) .$$

Example 3.7.6: rotation

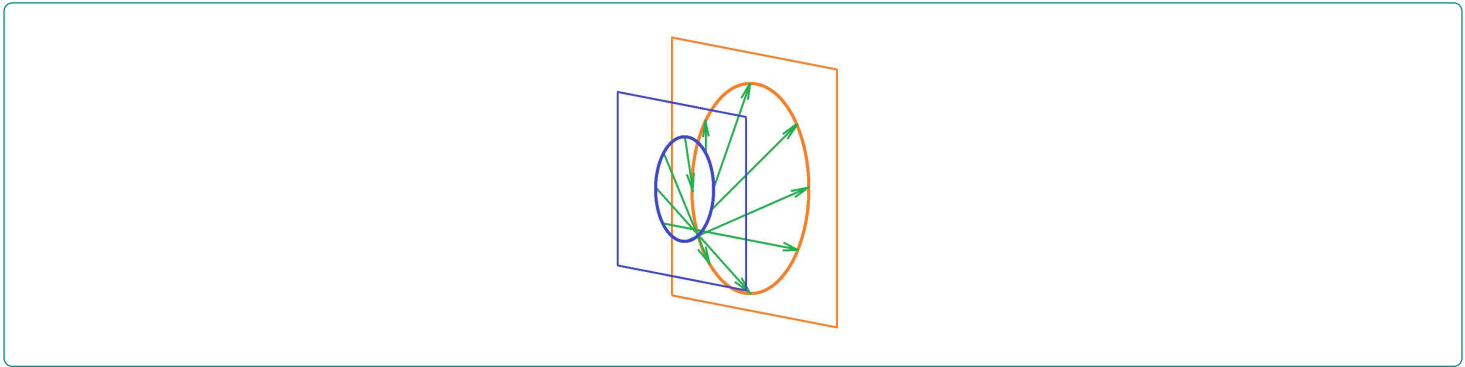
A *rotation* is carried out via a multiplication by a fixed complex number:

$$F(z) = z_0z .$$

Specifically, it has to be a number with modulus equal to 1:

$$z_0 = \cos \alpha + i \sin \alpha .$$

Meanwhile, with $z_0 = 2i$, we have the 90 degrees rotation with a stretch with a factor of 2:



For any complex function, we represent both the independent variable and the dependent variable in terms of their real and imaginary parts, just as vector functions. First:

$$x = u + iv ,$$

where u, v are real numbers. Second:

$$z = F(x) = f(u, v) + ig(u, v) ,$$

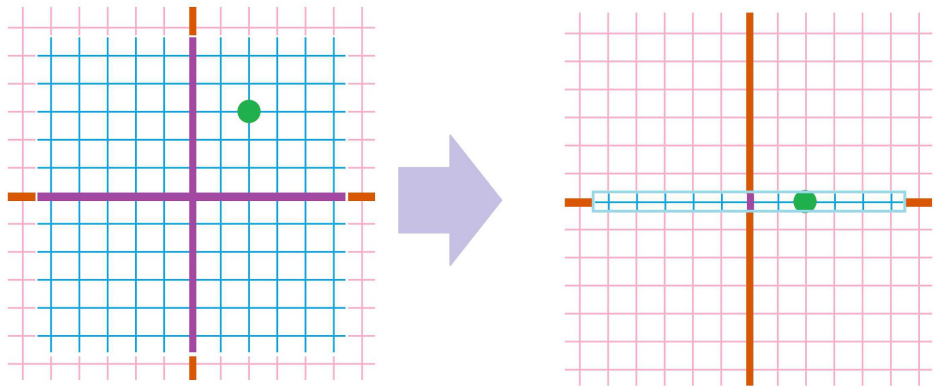
$f(u, v), g(u, v)$ are real-valued functions of two variables.

The two component functions, f and g , can be plotted. And so can the argument $\arg F$ and the module $||F||$ of the function.

Example 3.7.7: projections

The projections on the x - and y -axes are these:

$$F(z) = \operatorname{Re} z \quad \text{and} \quad F(z) = \operatorname{Im} z .$$

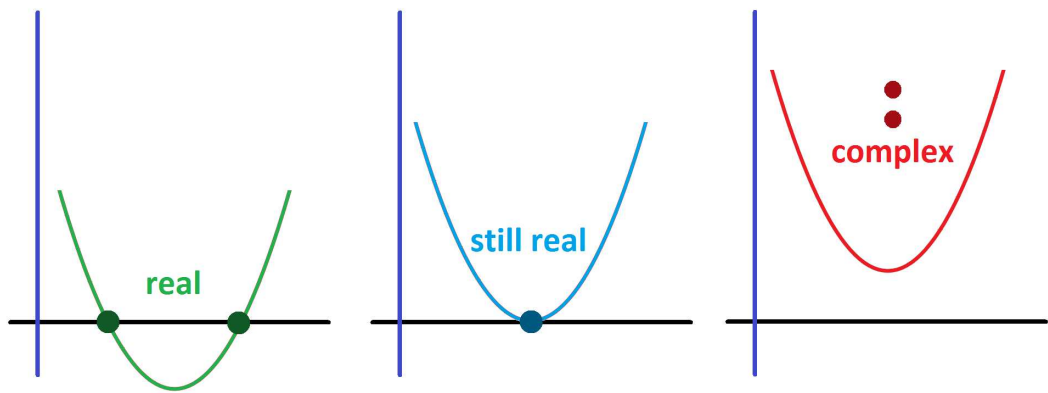


Example 3.7.8: quadratic

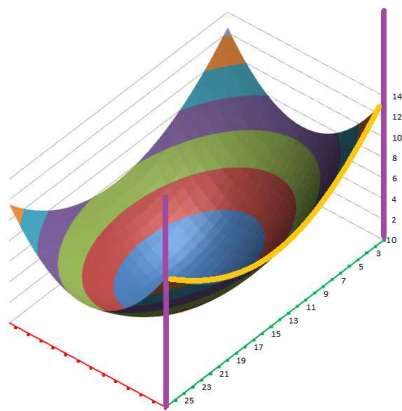
Let's consider a quadratic polynomial again:

$$f(x) = x^2 + px + q .$$

Recall that increasing the value of q makes the graph of $y = f(x)$ shift upward: its two x -intercepts start to get closer to each other, then merge, and finally disappear:



Note that an identical result is seen in a seemingly different situation. Suppose we have a paraboloid, then it produces a parabola as it is cut by a vertical plane. Suppose the paraboloid is moving horizontally. If it is fading away, the parabola is moving upward.



This illustrates what happens when our quadratic polynomial is seen as a function of a complex variable. We are plotting the real part of the function.

Warning!

Visualizing $F(x)$ via those functions, or as a vector field, may be misleading.

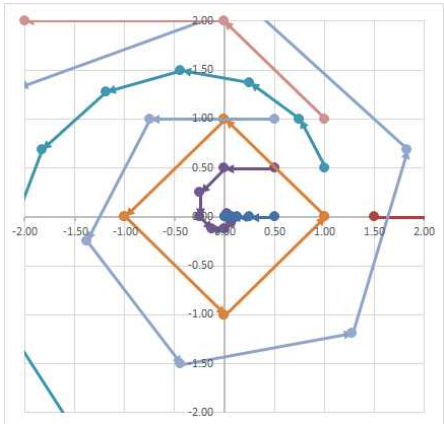
Complex functions are transformations of the complex plane!

Example 3.7.9: complex powers

This is a visualization of the power function over complex numbers. For several values of z , the values

$$z, z^2, z^3, \dots$$

are plotted as sequences.



One can see how the real part of z makes the multiplication by z stretch or shrink the number while

the imaginary part of x is responsible for rotating the number around 0. A special, square path is produced by $z = i$.

The definition is a copy of the one from [Chapter 2DC-2](#).

Definition 3.7.10: limit of a function

The *limit of a function* $z = F(x)$ at a point $x = a$ is defined to be the limit

$$\lim_{n \rightarrow \infty} F(x_n)$$

considered for all sequences $\{x_n\}$ within the domain of F excluding a that converge to a ,

$$a \neq x_n \rightarrow a \text{ as } n \rightarrow \infty,$$

when all these limits exist and are equal to each other. In that case, we use the notation:

$$\lim_{x \rightarrow a} F(x).$$

Otherwise, *the limit does not exist*.

Theorem 3.7.11: Locality

Suppose two functions f and g coincide in the vicinity of point a :

$$f(x) = g(x) \text{ for all } x \text{ with } ||x - a|| < \varepsilon,$$

for some $\varepsilon > 0$. Then, their limits at a coincide too:

$$\lim_{x \rightarrow a} f(x) = \lim_{x \rightarrow a} g(x).$$

Limits under algebraic operations... We will use the algebraic properties of the limits of sequences to prove virtually identical facts about limits of functions.

Let's re-write the main algebraic properties using the alternative notation.

Theorem 3.7.12: Algebra of Limits of Sequences

Suppose $a_n \rightarrow a$ and $b_n \rightarrow b$. Then

SR:	$a_n + b_n \rightarrow a + b$	CMR:	$c \cdot a_n \rightarrow ca$	for any complex c
PR:	$a_n \cdot b_n \rightarrow ab$	QR:	$a_n/b_n \rightarrow a/b$	provided $b \neq 0$

Each property is matched by its analog for functions.

Theorem 3.7.13: Algebra of Limits of Functions

Suppose $f(x) \rightarrow F$ and $g(x) \rightarrow G$ as $x \rightarrow a$. Then

SR:	$f(x) + g(x) \rightarrow F + G$	CMR:	$c \cdot f(x) \rightarrow cF$	for any complex c
PR:	$f(x) \cdot g(x) \rightarrow FG$	QR:	$f(x)/g(x) \rightarrow F/G$	provided $G \neq 0$

Just as before, the next concept is continuity:

Definition 3.7.14: continuous function

A function f is called *continuous at point a* if:

- The function $f(x)$ is defined at $x = a$.
- The limit of f exists at a .
- The two are equal to each other:

$$\lim_{x \rightarrow a} f(x) = f(a) .$$

Thus, the limits of continuous functions can be found by *substitution*.

Equivalently, a function f is continuous at a if

$$\lim_{n \rightarrow \infty} f(x_n) = f(a) ,$$

for any sequence $x_n \rightarrow a$.

A typical function we encounter is continuous at every point of its domain. The most important class of continuous functions is the following.

Theorem 3.7.15: Continuity of Polynomials

Every polynomial is continuous at every point.

Unlike vector functions, complex functions have more operations to worry about. The theorem follows from the following algebraic result.

Theorem 3.7.16: Algebra of Continuity

Suppose f and g are continuous at $x = a$. Then so are the following functions:

SR:	$f + g$	CMR:	$c \cdot f$	for any complex c
PR:	$f \cdot g$	QR:	f/g	provided $g(a) \neq 0$

3.8. Complex linear operators

Let’s consider linear operators again.

Suppose A is a linear operator with a 2×2 matrix with *real* entries. Suppose D is the discriminant of the characteristic polynomial of A . When $D > 0$, we have two distinct real roots covered by the *Classification Theorem of Linear Operators* with real eigenvalues. We also saw the transitional case when $D = 0$. Thus, we transition to the case $D < 0$ when the eigenvalues – as roots of the characteristic polynomial – are *complex*. We already know that complex numbers are just as good (or better!) than the real, so why not include this possibility?

Example 3.8.1: rotation

This is how we find the characteristic polynomial for the rotation and find the eigenvalues by solving it:

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \implies \det \left(\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) = \det \begin{bmatrix} -\lambda & -1 \\ 1 & -\lambda \end{bmatrix} = \lambda^2 + 1 = 0 \implies \lambda = \pm i .$$

The eigenvalues are imaginary! Let’s notice though that no *real* vector multiplied by an imaginary number can produce a real vector... Indeed, let’s try to find the eigenvectors; solve the matrix equation

$AV = \lambda V$ for V in \mathbf{R}^2 . In other words, we need to find real(!) x, y that satisfy:

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = i \begin{bmatrix} x \\ y \end{bmatrix} \implies \begin{cases} -y &= ix \\ x &= iy \end{cases}$$

Unless both zero, x and y can't be both real...

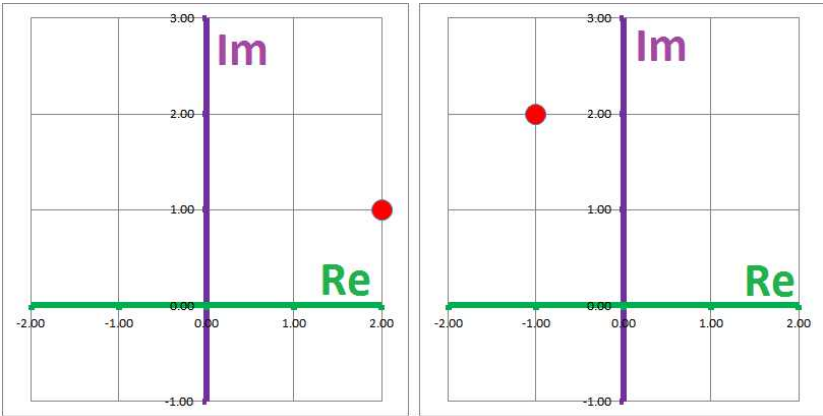
So, when the eigenvalues aren't real, there are no real eigenvectors.

But why stop here? Why not have all the numbers *and* vectors *and* matrices complex?

Just as a real 2-vector is a pair of real numbers, a *complex 2-vector* is a pair of complex numbers; for example,

$$V = \begin{bmatrix} 2 & + & i \\ -1 & + & 2i \end{bmatrix}.$$

This representation is illustrated below via the real and imaginary parts of either of the two components:

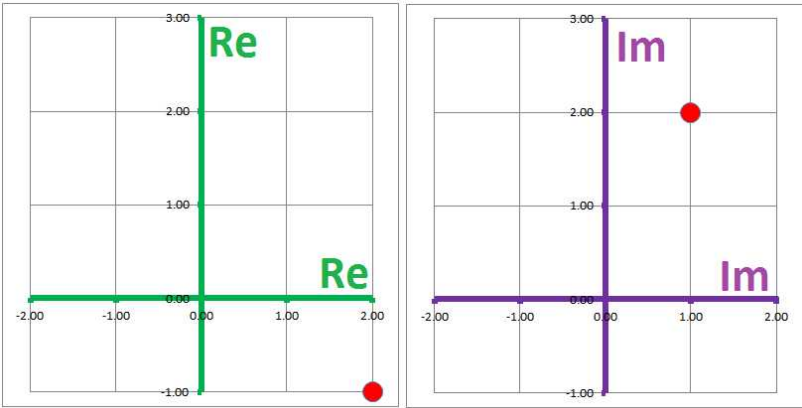


Here, we see the complex plane \mathbf{C} for the first and then the complex plane \mathbf{C} for the second component of the vector V . This is why we denote the set of all complex 2-vectors by \mathbf{C}^2 .

Furthermore, this vector V can be rewritten in terms of its *real and imaginary parts*:

$$V = \begin{bmatrix} 2 & + & i \\ -1 & + & 2i \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix} + i \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Each of these is a *real* vector and they are illustrated accordingly:



The former is of the main interest and it is located in the familiar real plane \mathbf{R}^2 .

Next, a *complex 2×2 matrix* A is simply a 2×2 table of complex numbers; for example:

$$A = \begin{bmatrix} 0 & 1 + i \\ i & 2 - 3i \end{bmatrix}.$$

The algebra of complex numbers – addition and multiplication – presented above allows us to carry out the operations – addition, scalar multiplication, and matrix multiplication – on these vectors and these matrices! Then, a complex matrix A defines a function:

$$A : \mathbb{C}^2 \rightarrow \mathbb{C}^2,$$

through matrix multiplication: $A(X) = AX$. Moreover, this function is a *linear operator* in the following sense:

$$A(\alpha X + \beta Y) = \alpha A(X) + \beta A(Y),$$

for any complex numbers α and β and any complex vectors X and Y .

But how can we use these functions to understand *real* linear operators? Its inputs and outputs are complex vectors and they have real parts as discussed above. So, we can restrict the domain of a complex linear operator to the real plane first and then take the real part of the output. The result is a familiar real linear operator:

$$B : \mathbb{R}^2 \rightarrow \mathbb{R}^2,$$

the *real part* of the complex linear operator A .

Let’s review our theory generalized this way. There is no difference:

Definition 3.8.2: determinant

The *determinant* of a complex 2×2 matrix is defined to be the following complex number:

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$$

Theorem 3.8.3: Non-zero Determinant

Suppose A is a complex 2×2 matrix. Then, $\det A \neq 0$ if and only if the solution set of the matrix equation $AX = 0$ consists of only 0 .

Definition 3.8.4: eigenvalue

Given a linear operator $A : \mathbb{C}^2 \rightarrow \mathbb{C}^2$, a (complex) number λ is called an *eigenvalue* of A if

$$A(V) = \lambda V$$

for some non-zero vector V in \mathbb{C}^2 . Then, V is called an *eigenvector* of A corresponding to λ .

Definition 3.8.5: eigenspace

For a (complex) eigenvalue λ of A , the *eigenspace* of a complex linear operator A corresponding to λ is defined and denoted by the following set in \mathbb{C}^2 :

$$E(\lambda) = \{V : A(V) = \lambda V\}$$

It is a very important fact that all the computations that we have performed on real matrices and vectors remain valid! Among the results that remain valid is the *Classification of Linear Operators* with real eigenvalues.

Linear operators represented by *real* matrices however will remain our exclusive interest.

Every linear operator represented by a real matrix A is still just a special case of a complex operator $A : \mathbb{C}^2 \rightarrow \mathbb{C}^2$. In fact, it will always have some complex (non-real) vectors among its values, unless $A = 0$. Its characteristic polynomial has *real* coefficients and, therefore, the *Classification Theorem of Quadratic Polynomials* presented in this chapter applies, as follows.

Theorem 3.8.6: Classification Theorem of Eigenvalues

Suppose A is a linear operator represented by a real 2×2 matrix A and D is the discriminant of its characteristic polynomial. Then the eigenvalues λ_1, λ_2 of A fall into one of the following three categories:

1. If $D > 0$, then λ_1, λ_2 are distinct real.

2. If $D = 0$, then λ_1, λ_2 are equal real.

3. If $D < 0$, then λ_1, λ_2 are complex conjugate.

We only need to address the last case.

3.9. Linear operators with complex eigenvalues

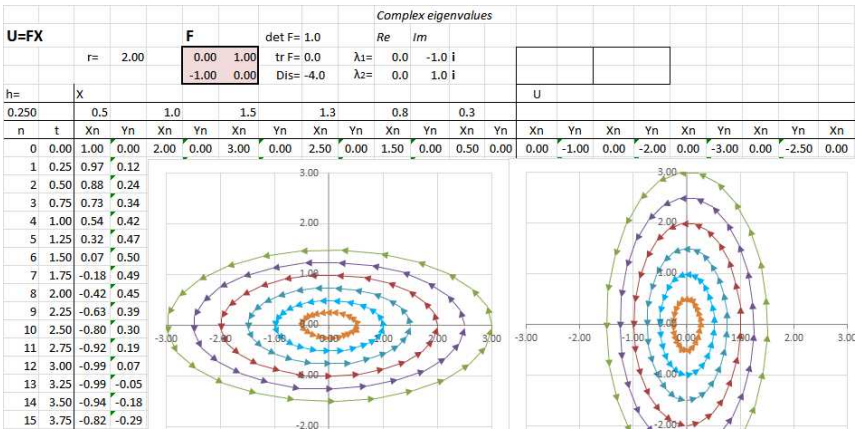
All numbers below are complex unless stated otherwise.

The first thing we notice about the case when the eigenvalues of a linear operator aren't real is that *there are no real eigenvectors*. There will be no eigenspaces shown in the *real* plane shown in the examples below.

Example 3.9.1: rotations

Consider a rotation through 90 degrees counterclockwise again:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \implies \lambda_{1,2} = \pm i .$$



To find the first eigenvector, we solve:

$$FV_1 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = i \begin{bmatrix} x \\ y \end{bmatrix} .$$

This gives as the following system of linear equations:

$$\begin{cases} -y = ix, \text{ AND} \\ x = iy \end{cases} \implies y = -ix .$$

We choose a *complex* eigenvector:

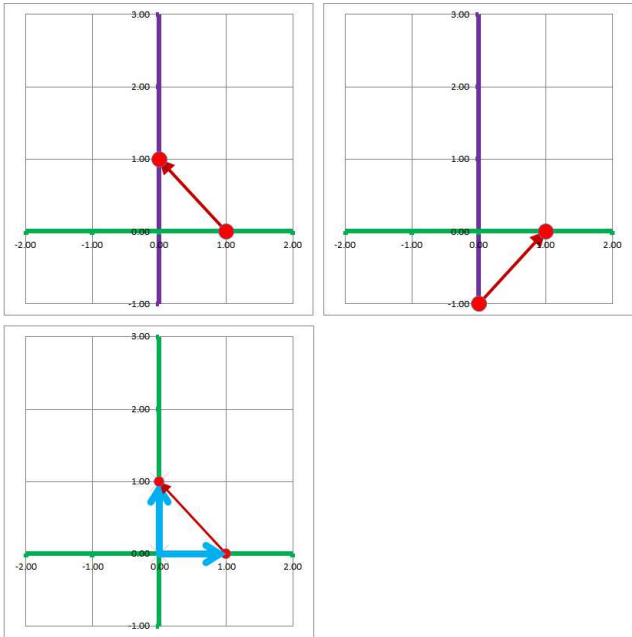
$$V_1 = \begin{bmatrix} 1 \\ -i \end{bmatrix} ,$$

and similarly an eigenvector for the second eigenvalue:

$$V_2 = \begin{bmatrix} 1 \\ i \end{bmatrix} .$$

The rest are (complex) multiples of these.

So, under F , complex vector V_k is multiplied by λ_k , and so is every of its multiples, $k = 1, 2$. Since these multiples are complex, this multiplication rotates (the vector of the geometric representation of) either component of this vector. Furthermore, the real part of this vector is also rotated – on the real plane. It is this rotation that we are interested in. It is shown in the second row below:

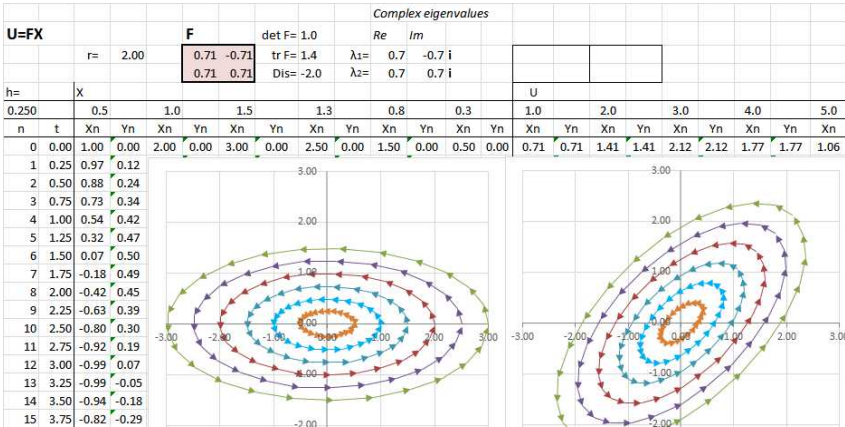


The general value is a linear combination – over the complex numbers – of our two eigenvectors:

$$X = \alpha V_1 + \beta V_2 \implies FX = \alpha i \begin{bmatrix} 1 \\ -i \end{bmatrix} + \beta (-i) \begin{bmatrix} 1 \\ i \end{bmatrix} .$$

Let's consider a *rotation through an arbitrary angle θ* :

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} .$$



Then, we have

$$\chi_A(\lambda) = (\cos \theta - \lambda)^2 + \sin^2 \theta = \cos^2 \theta - 2 \cos \theta \lambda + \lambda^2 + \sin^2 \theta = \lambda^2 - 2 \cos \theta \lambda + 1 .$$

Therefore,

$$\lambda_{1,2} = \frac{2 \cos \theta \pm \sqrt{(2 \cos \theta)^2 - 4}}{2} = \frac{2 \cos \theta \pm 2 \sqrt{\cos^2 \theta - 1}}{2} = \cos \theta \pm \sqrt{-\sin^2 \theta} = \cos \theta \pm i \sin \theta .$$

So, the argument of the complex eigenvalues is equal to the angle of rotation (up to a sign):

$$|\arg \lambda_{1,2}| = |\theta|.$$

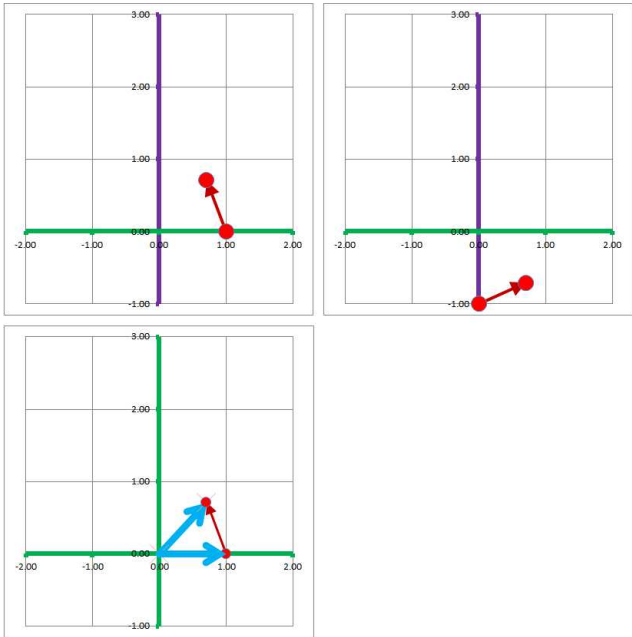
The eigenvectors are the same as in the last example:

$$V_1 = \begin{bmatrix} 1 \\ -i \end{bmatrix} \quad \text{and} \quad V_2 = \begin{bmatrix} 1 \\ i \end{bmatrix}.$$

Indeed, they are rotated through multiplication by the eigenvalues:

$$\lambda_k V_k = (\cos \theta \pm i \sin \theta) \begin{bmatrix} 1 \\ \mp i \end{bmatrix} = \begin{bmatrix} \cos \theta \pm i \sin \theta \\ \sin \theta \mp i \cos \theta \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} 1 \\ \mp i \end{bmatrix}.$$

Once again, we see how the real part of a complex eigenvector is rotated via complex multiplication:



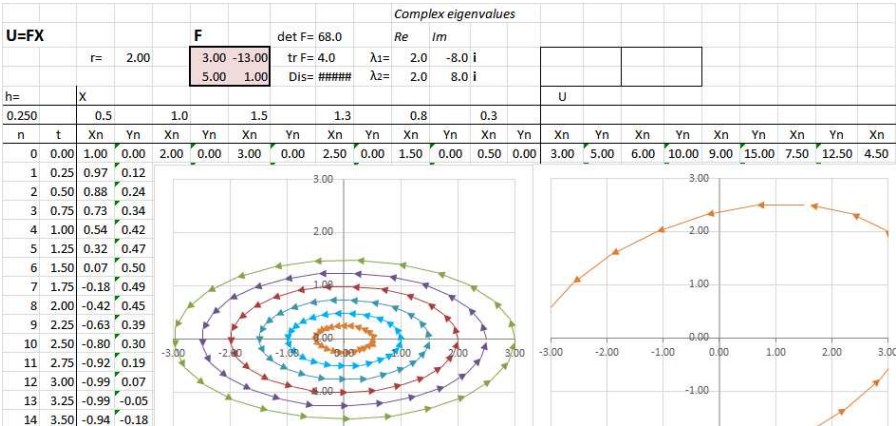
Warning!

This does not apply to vectors in \mathbf{C}^2 that aren't eigenvectors.

Example 3.9.2: rotation with stretch-shrink

Let's consider this linear operator:

$$\begin{cases} u = 3x - 13y \\ v = 5x + y \end{cases} \quad \text{and} \quad F = \begin{bmatrix} 3 & -13 \\ 5 & 1 \end{bmatrix}.$$



Our analysis starts with the characteristic polynomial:

$$\chi(\lambda) = \det(F - \lambda I) = \det \begin{bmatrix} 3 - \lambda & -13 \\ 5 & 1 - \lambda \end{bmatrix} = \lambda^2 - 4\lambda + 68.$$

We find the eigenvalues from the *Quadratic Formula*:

$$\lambda_{1,2} = 2 \pm 8i.$$

Now we find the eigenvectors. We solve the two equations:

$$FV_k = \lambda_k V_k, \quad k = 1, 2.$$

The first:

$$FV_1 = \begin{bmatrix} 3 & -13 \\ 5 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = (2 + 8i) \begin{bmatrix} x \\ y \end{bmatrix}.$$

This gives as the following system of linear equations:

$$\begin{cases} 3x - 13y = (2 + 8i)x, & \text{AND} \\ 5x + y = (2 + 8i)y \end{cases} \implies \begin{cases} (1 - 8i)x - 13y = 0 \\ 5x + (-1 - 8i)y = 0 \end{cases} \implies x = \frac{(1 + 8i)}{5}y.$$

We choose the first eigenvector to be:

$$V_1 = \begin{bmatrix} 1 + 8i \\ 5 \end{bmatrix}.$$

The second eigenvalue satisfies:

$$FV_2 = \begin{bmatrix} 3 & -13 \\ 5 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = (2 - 8i) \begin{bmatrix} x \\ y \end{bmatrix}.$$

We have the following system:

$$\begin{cases} 3x - 13y = (2 - 8i)x \\ 5x + y = (2 - 8i)y \end{cases} \implies \begin{cases} (1 + 8i)x - 13y = 0 \\ 5x + (-1 + 8i)y = 0 \end{cases} \implies x = \frac{(1 - 8i)}{5}y.$$

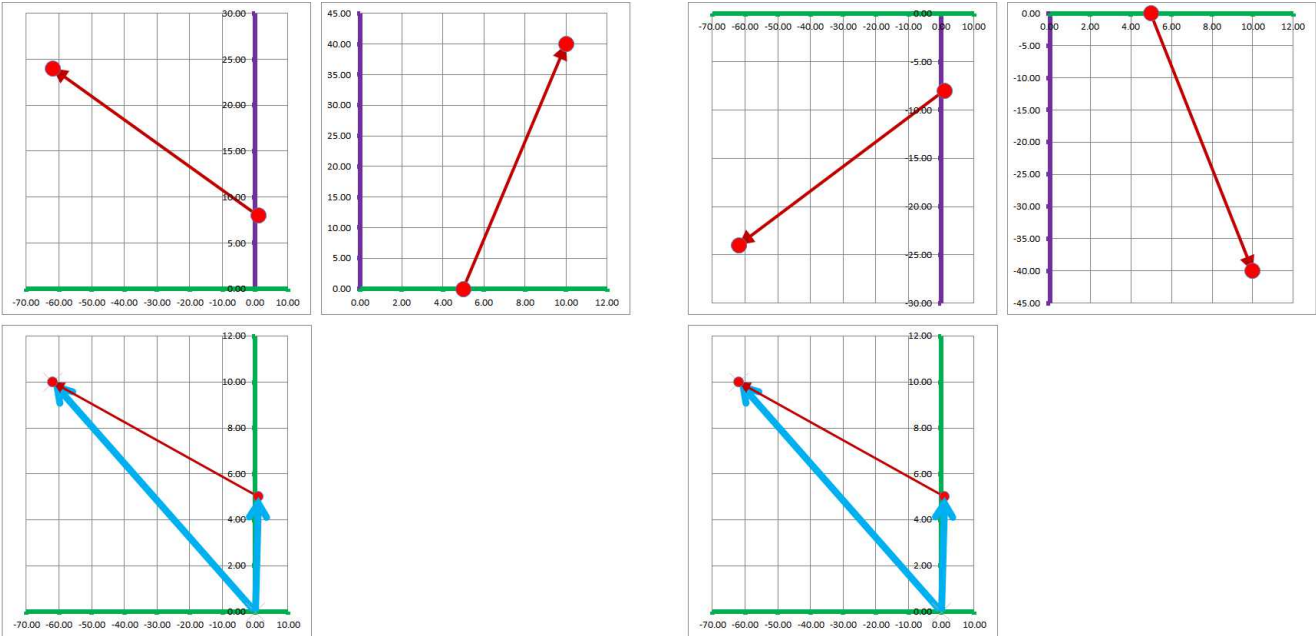
We choose the second eigenvector to be:

$$V_2 = \begin{bmatrix} 1 - 8i \\ 5 \end{bmatrix}.$$

The general complex value is a linear combination of the two:

$$X = \alpha V_1 + \beta V_2 \implies FX = \alpha(2 + 8i) \begin{bmatrix} 1 + 8i \\ 5 \end{bmatrix} + \beta(2 - 8i) \begin{bmatrix} 1 - 8i \\ 5 \end{bmatrix}.$$

We know the effect of F on these two vectors: they are rotated and stretched.



As you can see, stretching is the same for both components and both basis vectors. That is why we have – in addition to the rotation – a uniform stretch (re-scaling) for the real plane. This is shown in the second row.

This is the algebra for the above illustration:

$$\lambda_1 V_1 = (2 + 8i) \begin{bmatrix} 1 + 8i \\ 5 \end{bmatrix} = \begin{bmatrix} -62 + 24i \\ 10 + 40i \end{bmatrix} \quad \text{and} \quad \lambda_2 V_2 = (2 - 8i) \begin{bmatrix} 1 - 8i \\ 5 \end{bmatrix} = \begin{bmatrix} -62 - 24i \\ 10 - 40i \end{bmatrix} .$$

According to the *Classification Theorem of Quadratic Polynomials*, when the discriminant $D < 0$, the two roots are conjugate:

$$\lambda_{1,2} = a \pm bi .$$

They have the same modulus:

$$||\lambda_{1,2}|| = \sqrt{a^2 + b^2} .$$

This is why multiplying any complex number by either of the two numbers will produce the same rate of stretch. We conclude that \mathbf{C}^2 and, especially, the real plane \mathbf{R}^2 are stretched *uniformly*! This is the summary of our analysis.

Theorem 3.9.3: Classification of Linear Operators with Complex Eigenvalues

Suppose a real matrix F has two complex conjugate eigenvalues λ_1 and λ_2 . Then, the operator $U = FX$ does the following:

- It rotates the real plane through the angle θ that satisfies the following:

$$|\sin \theta| = |\arg \lambda_1| = |\arg \lambda_2| .$$

- It stretches-shrinks the plane uniformly by the following factor:

$$s = ||\lambda_1|| = ||\lambda_2|| .$$

We can also recast this theorem in exclusively “real” terms.

We will need to the following result:

Theorem 3.9.4: Non-positive Discriminant

If a quadratic polynomial,

$$x^2 + px + q ,$$

has a non-positive discriminant,

$$D = p^2 - 4q \leq 0 ,$$

then either of its roots,

$$x = \frac{1}{2}(-p \pm \sqrt{D}) ,$$

has its argument satisfy:

$$\sin \left(\arg x \right) = \frac{1}{2} \sqrt{4 - \frac{p^2}{q}} ,$$

and its module satisfy:

$$||x||^2 = q .$$

Proof.

The modulus of the roots is the following:

$$||x||^2 = (\operatorname{Re} x)^2 + (\operatorname{Im} x)^2 = \frac{1}{4}(p^2 + |D|) = \frac{1}{4}(p^2 - D) = \frac{1}{4}(p^2 - (p^2 - 4q)) = q.$$

And their argument satisfies the following:

$$\sin(\arg x) = \frac{\operatorname{Im} x}{||x||} = \frac{\frac{1}{2}\sqrt{D}}{\sqrt{q}} = \frac{1}{2}\sqrt{\frac{-D}{q}} = \frac{1}{2}\sqrt{-\frac{p^2 - 4q}{q}} = \frac{1}{2}\sqrt{4 - \frac{p^2}{q}}.$$

We defined the *trace* of a matrix as the sum of its diagonal elements:

$$\operatorname{tr} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = a + d.$$

Then the characteristic polynomial takes this form:

$$\begin{aligned} \chi(\lambda) &= \det \begin{bmatrix} a - \lambda & b \\ c & d - \lambda \end{bmatrix} \\ &= ad - a\lambda - \lambda d + \lambda^2 - bc \\ &= \lambda^2 - (a + d)\lambda + (ad - bc) \\ &= \lambda^2 - \operatorname{tr} F \lambda + \det F. \end{aligned}$$

We match this to the theorem above:

$$p = -\operatorname{tr} F, \quad q = \det F, \quad D = (\operatorname{tr} F)^2 - 4 \det F.$$

The roots are

$$\lambda_{1,2} = \frac{1}{2}(\operatorname{tr} F \pm \sqrt{D}).$$

When the roots are complex, the modulus is the following:

$$||\lambda_{1,2}|| = \sqrt{\det F}.$$

And their argument satisfies the following:

$$\sin(\arg \lambda_{1,2}) = \frac{1}{2}\sqrt{\frac{(\operatorname{tr} F)^2}{\det F} - 4}.$$

This is the final result.

Corollary 3.9.5: Non-positive Discriminant

Suppose a real matrix F satisfies:

$$D = (\operatorname{tr} F)^2 - 4 \det F \leq 0.$$

Then, the operator $U = FX$ does the following:

- It rotates the real plane through the following angle:

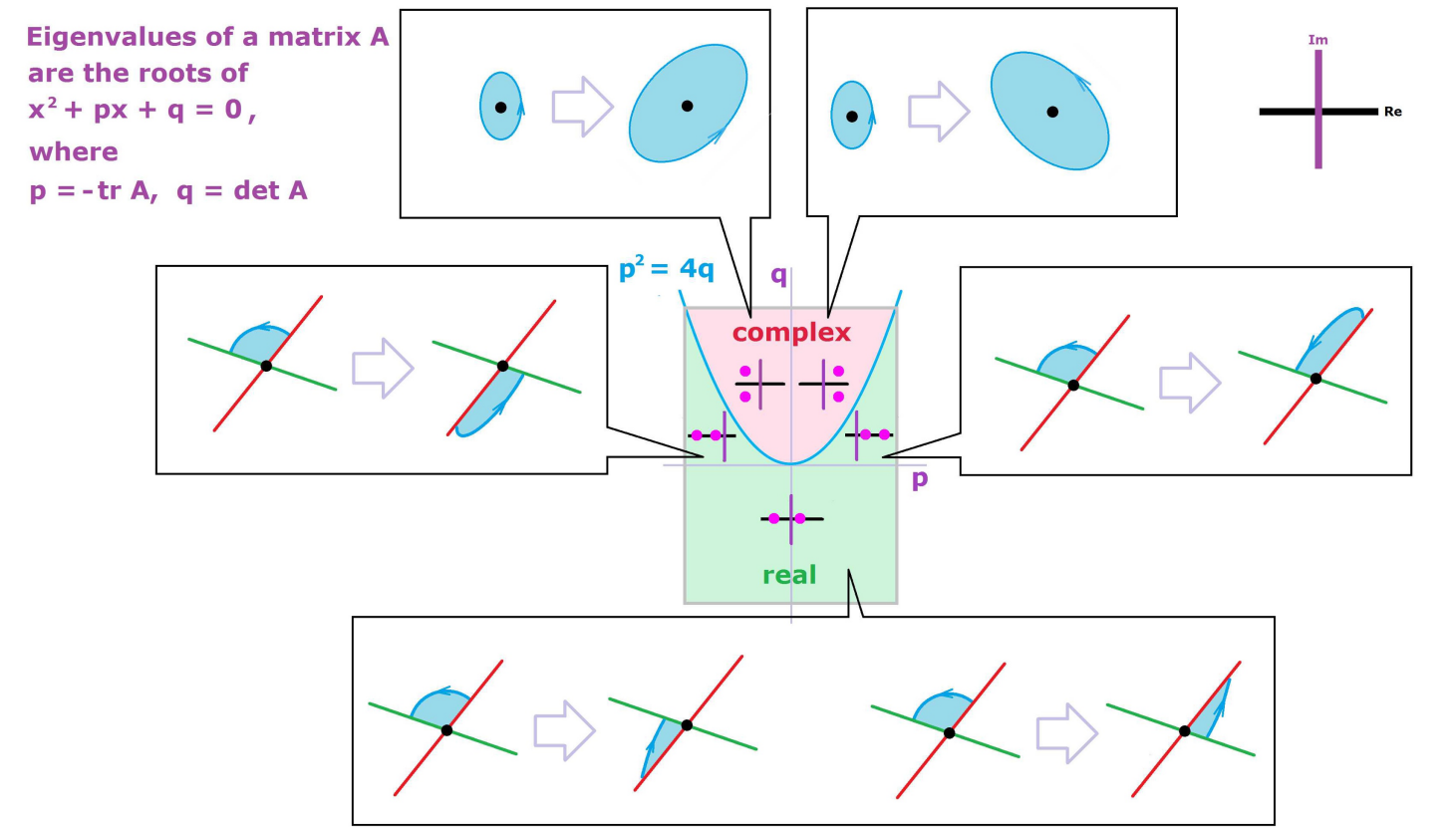
$$\theta = \sin^{-1} \left(\frac{1}{2} \sqrt{\frac{4 - (\operatorname{tr} F)^2}{\det F}} \right).$$

- It stretches-shrinks the plane uniformly by the following factor:

$$s = \sqrt{\det F}.$$

Yes, we have included the transitional case $D = 0$! Indeed, it has the same stretch but no rotation.

We finally put together our two classification theorems: the behavior of linear operators in terms of the eigenvalues – real or complex – of their matrices. We illustrate this classification below in the context of the *Classification of Roots of Quadratic Polynomials*:



3.10. Complex calculus

Even though complex numbers are represented by plane vectors, the formula for the derivative of a complex function doesn't follow the idea of the gradient. It rather follows, and is identical to, the definition of the derivative of a usual *numerical* function. We rely on the fact that the algebra is the same even though the nature of the numbers is different:

Definition 3.10.1: derivative of a complex function

The *derivative* of a complex function $u = f(x)$ at $x = a$ is defined to be the limit of the difference quotients at $x = a$ as the increment Δx is approaching 0, denoted by:

$$f'(x) = \lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a} = \lim_{||x-a|| \rightarrow 0} \frac{f(x) - f(a)}{x - a}$$

in that case the function is called *differentiable at $x = a$* .

This formula is made possible by the availability of multiplication and division.

Warning!

The formula is not the same as this:

$$\lim_{x \rightarrow a} \frac{f(x) - f(a)}{||x - a||}.$$

Warning!

The idea of the derivative isn't about the slope, rise over the run, anymore.

The derivative is still the instantaneous rate of change of the output relative to the input. Many results are familiar such as the following:

Theorem 3.10.2: Diff => Cont

If a function is differentiable, it is also continuous.

Example 3.10.3: derivative

Computations work out in the exactly the same manner. Let's compute $f'(2)$ from the definition for

$$f(x) = -x^2 - x.$$

Definition:

$$f'(2) = \lim_{h \rightarrow 0} \frac{f(2 + h) - f(2)}{h}.$$

To compute the difference quotient, we need to substitute twice:

$$f(2 + h) = -(2 + h)^2 - (2 + h), \quad f(2) = -2^2 - 2.$$

Now, we substitute into the definition:

$$\begin{aligned} f'(2) &= \lim_{h \rightarrow 0} \frac{[-(2 + h)^2 - (2 + h)] - [-2^2 - 2]}{h} \\ &= \lim_{h \rightarrow 0} \frac{-4 - 4h - h^2 - 2 - h + 4 + 2}{h} \\ &= \lim_{h \rightarrow 0} \frac{-5h - h^2}{h} \\ &= \lim_{h \rightarrow 0} (-5 - h) \\ &= -5 - 0 \\ &= -5. \end{aligned}$$

Theorem 3.10.4: Integer Power Formula

$$(x^n)' = nx^{n-1}$$

Proof.

The proof relies entirely on the formula:

$$a^n - b^n = (a - b)(a^{n-1} + a^{n-2}b + \dots + ab^{n-2} + b^{n-1}).$$

Exercise 3.10.5

Finish the proof.

There is a counterpart for each rule of differentiation!

Theorem 3.10.6: Algebra of Derivatives

Wherever complex functions f and g are differentiable, we have the following:

SR: $(f + g)' = f' + g'$

PR: $(fg)' = f'g + fg'$

CMR: $(cf)' = cf'$

QR: $(f/g)' = \frac{f'g - fg'}{g^2}$

for any complex c

provided $g \neq 0$

We can differentiate any polynomial easily now.

Theorem 3.10.7: Derivative of Polynomial

For any positive integer n and any complex numbers a_0, \dots, a_n , we have the following:

$$\begin{aligned} & (a_n x^n + a_{n-1} x^{n-1} + a_{n-2} x^{n-2} + \dots + a_2 x^2 + a_1 x + a_0)' \\ = & a_n n x^{n-1} + a_{n-1} (n-1) x^{n-2} + a_{n-2} (n-2) x^{n-3} + \dots + a_2 2x + a_1 \end{aligned}$$

As far as applications are concerned, there is no such a relation as “less” or “more” among complex numbers! That’s why we don’t have to worry about:

- 1. monotonicity
- 2. extreme points
- 3. concavity, etc.

Just as before, reversing differentiation is called *integration* and the resulting functions are called *antiderivatives*; F is an antiderivative of f if $F' = f$.

There is a counterpart for each rule of integration!

Theorem 3.10.8: Algebra of Antiderivatives

Wherever f and g are integrable, we have the following:

SR: $\int (f + g) dx = \int f dx + \int g dx$

PR: $\int f dg = fg - \int g df$

CMR: $\int (cf) dx = c \int f dx$

LCR: $\int f(mx + b) dx = \frac{1}{m} \int f(t) dt \Big|_{t=mx+b}$

3.11. Series and power series

All the definitions and theorems continue to be identical or virtually identical to the ones in Volume 3 (Chapter 3IC-5).

Definition 3.11.1: sequence of sums, partial sums

For a given sequence $\{z_n\} = \{z_n : n = s, s + 1, s + 2, \dots\}$ of complex numbers, its *sequence of sums, or partial sums*, $\{p_n : n = s, s + 1, s + 2, \dots\}$ is a sequence defined by the following recursive formula:

$$p_s = z_s, \quad p_{n+1} = p_n + z_n, \quad n = s, s + 1, s + 2, \dots$$

Definition 3.11.2: sum of series

For a sequence $\{z_n\}$, the limit S of its sequence of partial sums $\{p_n\}$ is called by the *sum of the sequence* or, more commonly, the *sum of the series*, denoted by:

$$S = \sum_{i=s}^{\infty} z_i = \lim_{n \rightarrow \infty} \sum_{i=s}^n z_i$$

This limit might also be infinite.

From the *Uniqueness of Limit* we derive the following.

Theorem 3.11.3: Uniqueness of Sum

A series can have only one limit (finite or infinite); i.e., if a and b are sums of the same series, then $a = b$.

From the *Component-wise Convergence of Sequences* we derive the following.

Theorem 3.11.4: Component-wise Convergence of Series

A series of complex numbers $\sum_{i=s}^{\infty} z_i$ converges to a complex number z if and only if both the real and the imaginary parts of the sequence of (partial) sums of z_k converge to the real and the imaginary parts of z respectively; i.e.,

$$\sum_{i=s}^{\infty} z_i = z \iff \operatorname{Re} \sum_{i=s}^{\infty} z_i = \operatorname{Re}(z) \quad \text{and} \quad \operatorname{Im} \sum_{i=s}^{\infty} z_i = \operatorname{Im}(z).$$

We continue in the same order as with real-values series:

Definition 3.11.5: converges absolutely

For a sequence $\{z_n\}$, the series:

$$\sum_{i=\infty}^n z_i,$$

converges absolutely if the series of its moduli,

$$\sum_{i=\infty}^n ||z_i||,$$

converges.

Now, the algebra of series.

Just as before, we can *multiply a convergent series by a number term by term*.

Theorem 3.11.6: Constant Multiple Rule for Series

Suppose $\{s_n\}$ is a sequence. For any integer a and any complex c , we have:

$$\sum_a^{\infty}(c \cdot s_n) = c \cdot \sum_a^{\infty} s_n$$

provided the series converges.

Just as before, we can *add two convergent series term by term*.

Theorem 3.11.7: Sum Rule for Series

Suppose $\{s_n\}$ and $\{t_n\}$ are sequences. For any integer a , we have:

$$\sum_a^{\infty}(s_n + t_n) = \sum_a^{\infty} s_n + \sum_a^{\infty} t_n$$

provided the two series converge.

Exercise 3.11.8

Prove these theorems.

Definition 3.11.9: geometric series

The series produced by the geometric progression $a_n = a \cdot r^n$ with ratio r (a complex number)

$$\sum_{n=s}^{\infty} ar^n = ar^1 + ar^2 + ar^3 + ar^4 + \dots + ar^n + \dots$$

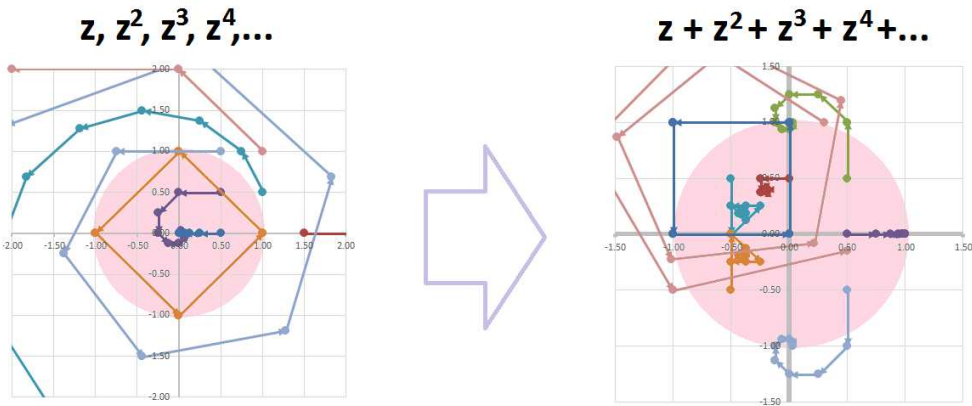
is called the *geometric series* with ratio r .

Example 3.11.10: complex power series

This is a visualization of a power series over complex numbers. We consider the series:

$$z + z^2 + z^3 + \dots$$

For several values of z , the sequence and then the partial sums of the series are plotted. We can see how starting from a point inside the disk $||z|| \leq 1$ produces divergence and outside produces divergence:



Indeed, the series

$$1 + z + z^2 + z^3 + \dots$$

is a geometric series with the ratio $r = z$. Therefore, it converges for all $\|z\| < 1$, according to the theorem, and diverges for all $\|z\| > 1$. In other words, it converges on the disk of radius 1 centered at 0 in \mathbf{C} and diverges outside of it. This circle is the domain of the function defined by the series. We even have a formula for this function:

$$1 + z + z^2 + z^3 + \dots = \frac{1}{1 - z}.$$

The difference between the two is in the *domain*.

Its sum is found the same way as for real variable. The only difference is that the absolute value is replaced with the modulus:

Theorem 3.11.11: Sum of Geometric Series

The geometric series with ratio r converges absolutely when $\|r\| < 1$ and diverges when $\|r\| > 1$. In the former case, the sum is:

$$\sum_{n=0}^{\infty} ar^n = \frac{a}{1 - r}.$$

Exercise 3.11.12

Prove the theorem.

Example 3.11.13: trig representation

There is a hint at the idea of complex power series in [Chapter 4HD-1](#). Let's compare the Taylor series of the sine, the cosine, and the exponential function. The sine is odd, and its Taylor series only includes odd terms:

$$\sin x = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k + 1)!} x^{2k+1}.$$

The cosine is even, and its Taylor series only includes even terms:

$$\cos x = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} x^{2k}.$$

If we write them one under the other, we see how they complement each other:

n	0	1	2	3	4	5	...
$\sin x$		x		$-\frac{x^3}{3!}$		$\frac{x^5}{5!}$...
$\cos x$	1		$-\frac{x^2}{2!}$		$\frac{x^4}{4!}$...

What if we add the exponential function to this?

n	0	1	2	3	4	5	...
$\sin x$	0	x	0	$-\frac{x^3}{3!}$	0	$\frac{x^5}{5!}$...
$\cos x$	1	0	$-\frac{x^2}{2!}$	0	$\frac{x^4}{4!}$	0	...
e^x	1	x	$\frac{x^2}{2!}$	$\frac{x^3}{3!}$	$\frac{x^4}{4!}$	$\frac{x^5}{5!}$...

This looks almost like addition!.. except for those minus signs. That's where i comes in. Let's substitute $x = it$:

$$\begin{aligned} e^{it} &= 1 + it + \frac{(it)^2}{2!} + \frac{(it)^3}{3!} + \frac{(it)^4}{4!} + \frac{(it)^5}{5!} + \dots \\ &= 1 + it - \frac{t^2}{2!} - \frac{it^3}{3!} + \frac{t^4}{4!} + \frac{it^5}{5!} - \dots \\ &= \left(1 - \frac{t^2}{2!} + \frac{t^4}{4!} - \dots\right) + i \left(t - \frac{t^3}{3!} + \frac{t^5}{5!} - \dots\right) \\ &= \cos t + i \sin t. \end{aligned}$$

It is called *Euler's formula*. More general is the following:

$$e^{a+bi} = e^a(\cos b + i \sin b).$$

In complex calculus, functions are more interrelated!

Exercise 3.11.14

Show that $\sin^2 x + \cos^2 x = 1$.

Definition 3.11.15: power series

A sequence $\{q_n\}$ of polynomials given by a recursive formula:

$$q_{k+1}(x) = q_k(x) + c_{k+1}(x - a)^{k+1}, \quad k = 0, 1, 2, \dots,$$

for some fixed (complex) number a and a sequence of (complex) numbers $\{c_k\}$, is called a *power series centered at a* . The function represented by the limit of q_n is called the *sum of the series*, written as:

$$f(x) = c_0 + c_1(x - a) + c_2(x - a)^2 + \dots = \sum_k c_k(x - a)^k = \lim_{k \rightarrow \infty} q_k(x)$$

for all x for which the limit exists.

Then the three power series in the example above may serve as the *definitions* of these three functions of complex variable:

n	0	1	2	3	4	5	...
$\sin x =$		x		$-\frac{x^3}{3!}$		$+\frac{x^5}{5!}$	$+ \dots$
$\cos x =$	1		$-\frac{x^2}{2!}$		$+\frac{x^4}{4!}$		$- \dots$
$e^x =$	1	$+x$	$+\frac{x^2}{2!}$	$+\frac{x^3}{3!}$	$+\frac{x^4}{4!}$	$+\frac{x^5}{5!}$	$+ \dots$

pending the proof of their convergence.

We will accept the following without proof.

Theorem 3.11.16: Weierstrass M-Test

Consider the power series

$$\sum_{n=0}^{\infty} c_n(z - a)^n.$$

Suppose there exists such a sequence of non-negative real numbers $\{M_n\}$ that

$$|c_n(z - a)^n| \leq M_n, \quad n = 0, 1, 2, \dots,$$

for all z in an open disk D around a and the series $\sum_{n=0}^{\infty} M_n$ (of real numbers) converges. Then the power series converges uniformly on D .

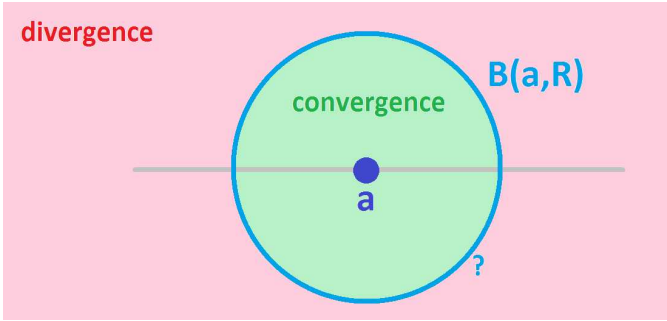
Convergence of the three series above follows.

Definition 3.11.17: radius of convergence

The greatest lower bound r (that could be infinite) of the distances from a to a point for which a power series centered at a diverges is called the *radius of convergence* of the series.

This definition is legitimate according to the *Existence of sup Theorem* from [Chapter 2DC-1](#).

We finally can see where the word “radius” comes from.



Theorem 3.11.18: Radius of Convergence

Suppose r is the radius of convergence of a power series

$$\sum_{n=0}^{\infty} c_n(x - a)^n.$$

Then we have:

1. When $r < \infty$, the domain of the series is a disk $B(a, R)$ in \mathbb{C} of radius r centered at a with some points on its boundary possibly included.
2. When $r = \infty$, the domain of the series is the whole \mathbb{C} .

Example 3.11.19: domains

A few basic series as functions:

series	sum	domain
$\sum_{k=0}^{\infty} x^k$	$= \frac{1}{1-x}$	$B(0,1)$
$\sum_{k=0}^{\infty} \frac{(-1)^k}{k!} x^k$	$= e^x$	\mathbf{C}
$\sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} x^{2k+1}$	$= \sin x$	\mathbf{C}
$\sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} x^{2k}$	$= \cos x$	\mathbf{C}

3.12. Solving ODEs with power series

Recall from Chapter 3IC-5 the following important results.

Theorem 3.12.1: Uniqueness of Power Series

If two power series are equal, as functions, on an open interval $(a-r, a+r)$, $r > 0$, then their corresponding coefficients are equal, i.e.,

$$\begin{aligned} \sum_{n=0}^{\infty} c_n(x-a)^n &= \sum_{n=0}^{\infty} d_n(x-a)^n \text{ for all } a-r < x < a+r \\ \implies c_n &= d_n \text{ for all } n = 0, 1, 2, 3, \dots \end{aligned}$$

Theorem 3.12.2: Term-by-Term Differentiation and Integration

Suppose $R > 0$ is the radius of convergence of a power series

$$f(x) = \sum_{n=0}^{\infty} a_n(x-a)^n.$$

Then the function f represented by this power series is differentiable (and, therefore, integrable) on the open disk $|x-a| < R$ and the power series representations of its derivative and its antiderivative converge on this disk and are found by term by term differentiation and integration of the power series of f , i.e.,

$$f'(x) = \left(\sum_{n=0}^{\infty} c_n(x-a)^n \right)' = \sum_{n=0}^{\infty} (c_n(x-a)^n)' = \sum_{n=1}^{\infty} n c_n(x-a)^{n-1},$$

and

$$\int f(x) \, dx = \int \left(\sum_{n=0}^{\infty} c_n(x-a)^n \right) dx = \sum_{n=0}^{\infty} \int c_n(x-a)^n \, dx = \sum_{n=0}^{\infty} \frac{c_n}{n+1} (x-a)^{n+1}.$$

Example 3.12.3: ODEs of first order

Suppose we need to solve this initial value problem (we pretend we don't know the answer):

$$y' = ky, \quad y(0) = y_0.$$

The solution is the same as the one presented in [Chapter 3IC-5](#). We assume that the unknown function $y = y(x)$ is differentiable and, therefore, is represented by a term-by-term differentiable power series. We differentiate the series and then match the terms according to the equation:

y	$=$	c_0	$+$	c_1x	$+$	c_2x^2	$+$	c_3x^3	$+$	\dots	$+$	c_nx^n	$+$	\dots
y'	$=$			c_1	$+$	$2c_2x$	$+$	$3c_3x^2$	$+$	\dots	$+$	nc_nx^{n-1}	$+$	\dots
\Rightarrow				\swarrow		\swarrow		\swarrow		\dots		\swarrow		
y'	$=$	c_1	$+$	$2c_2x$	$+$	$3c_3x^2$	$+$	\dots	$+$	nc_nx^{n-1}	$+$	$(n+1)c_{n+1}x^n$	$+$	\dots
\parallel		\parallel		\parallel		\parallel		\parallel		\parallel		\parallel		
$k \cdot y$	$=$	kc_0	$+$	kc_1x	$+$	kc_2x^2	$+$	\dots	$+$	$kc_{n-1}x^{n-1}$	$+$	kc_nx^n	$+$	\dots

According to the *Uniqueness of Power Series*, the coefficients have to match! Thus, we have a sequence of equations:

c_1	$2c_2$	$3c_3$	\dots	$(n+1)c_{n+1}$	$+\dots$
\parallel	\parallel	\parallel		\parallel	
kc_0	kc_1	kc_2	\dots	kc_n	$+\dots$

We can start solving these equations from left to right:

c_1	$\Rightarrow c_1 = kc_0$	$2c_2$	$\Rightarrow c_2 = k^2c_0/2$	$3c_3$	\dots
\parallel		\parallel		\parallel	
kc_0	$\Rightarrow kc_1 = k^2c_0$		$\Rightarrow kc_2 = k^3c_0/2$	\dots	

The condition $y(0) = y_0$ means that $c_0 = y_0$. Therefore,

$$c_n = y_0 \frac{k^n}{n!}.$$

We recognize the resulting series:

$$y = \sum_{n=0}^\infty y_0 \frac{k^n}{n!} x^n = y_0 \sum_{n=0}^\infty \frac{1}{n!} (kx)^n = y_0 e^{kx}.$$

Note to get the n th term we solve a system of linear equations with the following the augmented matrix:

$$\begin{bmatrix} k & -1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & k & -2 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & k & -3 & 0 & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots & k & -n & 0 \end{bmatrix}.$$

Warning!

Recognizing the resulting series isn't to be expected.

Exercise 3.12.4

Solve the initial value problem:
$$y' = ky + 1, \quad y(0) = y_0.$$

Example 3.12.5: ODEs of second order

Suppose we need to solve this initial value problem:

$$y'' = -y, \quad y(0) = y_0, \quad y'(0) = v_0.$$

Again, we assume that the unknown function $y = y(x)$ is represented by a term-by-term differentiable power series.

We differentiate the series twice and then match the terms according to the equation:

y
 y'
 y''
 \implies
 y''
 \parallel
 $-y$

$=$
 $=$
 $=$
 $=$
 $=$
 \parallel
 $=$

c_0
 c_1
 $$
 $$
 $2c_2$
 \parallel
 $-c_0$

$+$
 $+$
 $+$
 $+$
 $+$
 \parallel
 $+$

c_1x
 $$
 $$
 $$
 $2c_2x$
 \parallel
 $-c_1x$

$+$
 $+$
 $+$
 $+$
 $+$
 \parallel
 $+$

c_2x^2
 $2c_2x$
 $2c_2$
 $$
 $3 \cdot 2c_3x$
 \parallel
 $-c_2x^2$

$+$
 $+$
 $+$
 $+$
 $+$
 \parallel
 $+$

c_3x^3
 $3c_3x^2$
 $3 \cdot 2c_3x$
 $$
 $4 \cdot 3c_4x^2$
 \parallel
 $-c_3x^3$

$+$
 $+$
 $+$
 $+$
 $+$
 \parallel
 $+$

c_4x^4
 $4c_4x^3$
 $4 \cdot 3c_4x^2$
 $$
 $5 \cdot 4c_5x^3$
 \parallel
 $$

$+$
 $+$
 $+$
 $+$
 $+$
 \parallel
 $$

c_5x^5
 $5c_5x^4$
 $5 \cdot 4c_5x^3$
 $$
 \dots
 \parallel
 \dots

$+$
 $+$
 $+$
 $+$
 $+$
 \parallel
 \dots

The coefficients have to match:

$2c_2$
 \parallel
 $-c_0$

$3 \cdot 2c_3$
 \parallel
 $-c_1$

$4 \cdot 3c_4$
 \parallel
 $-c_2$

$5 \cdot 4c_5$
 \parallel
 $-c_3$

\dots
 \dots
 \dots

$n(n-1)c_n$
 \dots
 $-c_{n-2}$

\dots
 \dots
 \dots

We can start solving these equations from left to right, odd separate from even. First, for even n :

$2c_2$
 \parallel
 $-c_0$

$\implies c_2 = c_0/2$

$4 \cdot 3c_4$
 \parallel
 $\implies -c_2 = -c_0/2$

$\implies c_4 = -c_0/(4 \cdot 3 \cdot 2)$

$\dots \implies c_n = \pm c_0/n!$

\dots

The condition $y(0) = y_0$ means that $c_0 = y_0$. Therefore,

$n \text{ even} \implies c_n = (-1)^{n/2+1} \frac{y_0}{n!}.$

We recognize the resulting series:

$$\sum_{\text{even}, n=0}^{\infty} (-1)^{n/2+1} \frac{y_0}{n!} x^n = y_0 \sum_{\text{even } n=0}^{\infty} (-1)^{n/2+1} \frac{1}{n!} x^n = y_0 \cos x.$$

Second, for odd n :

$3 \cdot 2c_3$
 \parallel
 $-c_1$

$\Rightarrow c_3 = c_1/(3 \cdot 2)$

$5 \cdot 4c_5$
 \parallel
 $\Rightarrow -c_3 = -c_1/(3 \cdot 2)$

$\Rightarrow c_5 = -c_1/(5 \cdot 4 \cdot 3 \cdot 2)$

$\dots \Rightarrow c_n = \pm c_1/n!$

The condition $y'(0) = v_0$ means that $c_1 = v_0$. Therefore,

$n \text{ odd} \implies c_n = (-1)^{(n+1)/2} \frac{v_0}{n!}.$

We recognize the resulting series:

$$y = \sum_{\text{odd } n=1}^{\infty} (-1)^{(n+1)/2} \frac{v_0}{n!} x^n = v_0 \sum_{\text{odd } n=1}^{\infty} (-1)^{(n+1)/2} \frac{1}{n!} x^n = v_0 \sin x.$$

This is the result:

$$y = y_0 \cos x + v_0 \sin x.$$

Exercise 3.12.6

Provide the augmented matrix for this system of linear equations.

Exercise 3.12.7

Solve the initial value problem:

$$y'' + xy' + y = 0, \quad y(0) = 1, \quad y'(0) = 1.$$

When the resulting series isn't recognizable, it is used to approximate the answer via its partial sums. The accuracy of this approximation is given by the error bound for Taylor polynomials given in [Chapter 4HD-1](#), as follows.

Theorem 3.12.8: Error Bound

Suppose a function $y = y(x)$ is $(n + 1)$ times differentiable at $x = a$. Suppose also that for each $i = 0, 1, 2, \dots, n + 1$, we have

$$|y^{(i)}(t)| < K_i \text{ for all } t \text{ between } a \text{ and } x,$$

and some real number K_i . Then

$$e_n(x) = |y(x) - T_n(x)| \leq K_{n+1} \frac{|x - a|^{n+1}}{(n + 1)!},$$

where T_n is the n th Taylor polynomial of y .

Chapter 4: Systems of ODEs

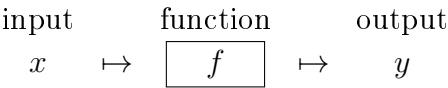
Contents

4.1 Parametric curves	254
4.2 The predator-prey model	263
4.3 Qualitative analysis of the predator-prey model	267
4.4 Solving the Lotka–Volterra equations	270
4.5 Vector fields and systems of ODEs	273
4.6 Discrete systems of ODEs	279
4.7 Qualitative analysis of systems of ODEs	283
4.8 The vector notation and linear systems	287
4.9 Classification of linear systems	293
4.10 Classification of linear systems, continued	298

4.1. Parametric curves

Functions process an input of any nature and produce an output of any nature.

In general, we represent a function diagrammatically as a *black box* that processes the input and produces the output:



Functions in multidimensional spaces take points or vectors as the input and produce points or vectors of various dimensions as the output. We can say that the input X is in \mathbf{R}^n and the output $U = F(X)$ of X is in \mathbf{R}^m :

$$F : \begin{array}{c} P \\ \text{in } \mathbf{R}^n \end{array} \mapsto \begin{array}{c} U \\ \text{in } \mathbf{R}^m \end{array}$$

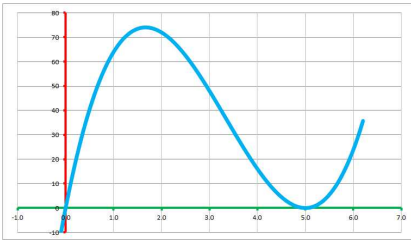
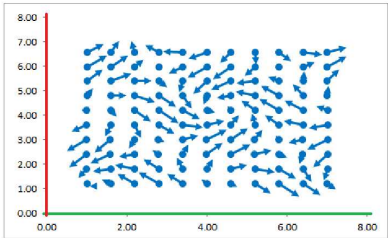
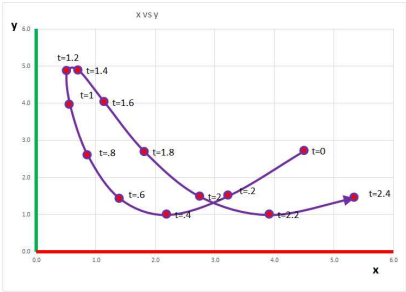
Then, the domain of such a function is in \mathbf{R}^n and the range (image) is in \mathbf{R}^m . The domain can be less than the whole space. In fact, the function can be defined on the nodes of a partition of the subset of \mathbf{R}^n . Below we illustrate the four possibilities for $n = 1, 2$ and $m = 1, 2$:

Functions illustrated

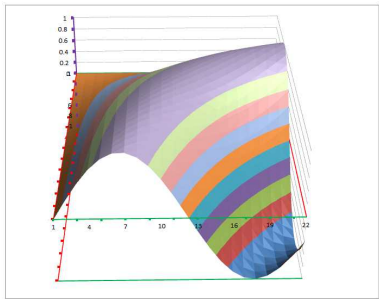
Codomain

dim = 2

dim = 1



dim = 1



dim = 2

Domain

In addition to the usual functions, we see

- a parametric curve $P : t \mapsto (x, y)$ or $P : t \mapsto \langle x, y \rangle$,
- a function of two variables $F : (x, y) \mapsto z$ or $F : \langle x, y \rangle \mapsto z$, and
- a vector field on the plane $V : (x, y) \mapsto \langle u, v \rangle$.

We need to learn how, instead of treating parametric curves one axis at a time, to study them as functions with multidimensional inputs and outputs. Vector algebra will be especially useful.

We will refer to as a *parametric curve* to

- any function of the real variable, i.e., the domain lies inside \mathbf{R} , and
- with its values in \mathbf{R}^m for some $m = 1, 2, 3, \dots$

In this section we will limit ourselves to the interpretation of these functions via *motion*. The independent variable is then *time* and the value is the *location*.

A *point* is the simplest curve. Such a curve with no motion is provided by a *constant function*.

A *straight line* is the second simplest curve.

We start with lines in \mathbf{R}^2 . We already know how to represent straight lines on the plane; the first method is the *slope-intercept form*:

$$y = mx + b.$$

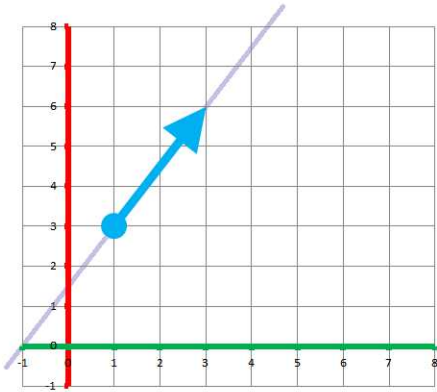
This method does not include vertical lines: the slope is infinite! In our study of curves (specifically to represent motion) on the plane, there are no preferred directions and then it is unacceptable to exclude any straight lines. The second method is *implicit*:

$$px + qy = r.$$

The case of $p \neq 0, q = 0$ gives us a vertical line. The third method is *parametric*. It has a dynamic interpretation.

Example 4.1.1: straight motion

Suppose we would like to trace the line that starts at the point $(1, 3)$ and proceeds in the direction of the vector $\langle 2, 3 \rangle$.

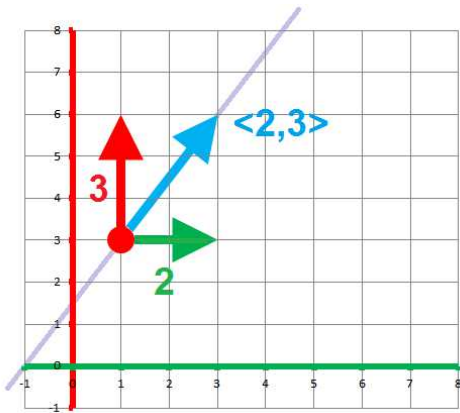


We use *motion* as a starting point and as well as a metaphor for parametric curves, as follows. We start moving ($t = 0$)

- from the point $P_0 = (1, 3)$,
- under a constant velocity of $V = \langle 2, 3 \rangle$.

We move

- 2 feet per second horizontally, and
- 3 feet per second vertically.



In terms of vectors, if we are at point P now, we will be at point $P + V$ after one second. For example, we are at $P_1 = P_0 + V = (1, 3) + \langle 2, 3 \rangle = (3, 6)$ at time $t = 1$. We then have already two points on our parametric curve P :

$$P(0) = P_0 = (1, 3) \quad \text{and} \quad P(1) = P_1 = (3, 6).$$

Of course, these are also the *values* of our function.

Let's find the formulas for this function. Early on, it's OK to do this component-wise. Then

$$P(t) = (x(t), y(t)),$$

and

$$x(0) = 1, \quad x(1) = 3 \quad \text{and} \quad y(0) = 3, \quad y(1) = 6.$$

These functions must be linear; therefore, we have:

$$x(t) = 1 + 2t \quad \text{and} \quad y(t) = 3 + 3t.$$

This is a parametric curve... but not an acceptable answer if we are to learn how to use vectors!

The four coefficients, of course, come from the specific numbers that give us P_0 and V . Let's assemble the two coordinate function into one parametric curve:

$$P(t) = (x(t), y(t)) = (1 + 2t, 3 + 3t).$$

This is still not good enough; we still can't see the P_0 and V directly! We continue by using vector algebra:

$$\begin{aligned} P(t) &= (1 + 2t, 3 + 3t) && \text{We use vector addition.} \\ &= (1, 3) + \langle 2t, 3t \rangle && \text{Then scalar multiplication.} \\ &= (1, 3) + t \langle 2, 3 \rangle && \text{And finally substitute.} \\ &= P_0 + tV. \end{aligned}$$

We have discovered a *vector* representation of straight motion:

$$P(t) = P_0 + tV,$$

where P_0 is the initial location and V is the (constant) velocity.

Warning!

One can, of course, move along a straight line at a *variable* velocity.

So,

position at time t = initial position + $t \cdot$ velocity

We stated this conclusion before! This time only the context has changed. The pattern is clear: the line starting at the point (a, b) in the direction of the vector $\langle u, v \rangle$ is represented parametrically as:

$$P(t) = (a, b) + t \langle u, v \rangle .$$

Similar for dimension 3: the line starting at the point (a, b, c) in the direction of the vector $\langle u, v, w \rangle$ is represented as:

$$P(t) = (a, b, c) + t \langle u, v, w \rangle .$$

And so on.

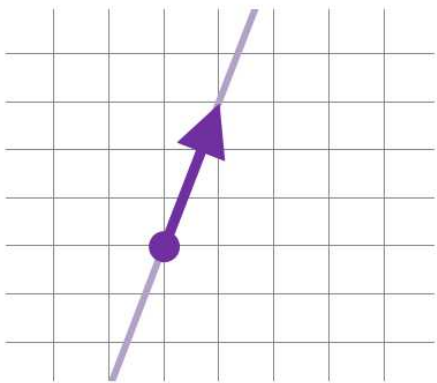
At the next level, we'd rather have no references to neither the dimension of the space nor the specific coordinates.

Definition 4.1.2: parametric curve of the uniform motion

Suppose P_0 is a point in \mathbf{R}^m and V is a vector. Then the *parametric curve of the uniform motion through P_0 with the initial velocity of V* is the following:

$$P(t) = P_0 + tV$$

Then, the *line through P_0 in the direction of V* is the path (image) of this parametric curve.



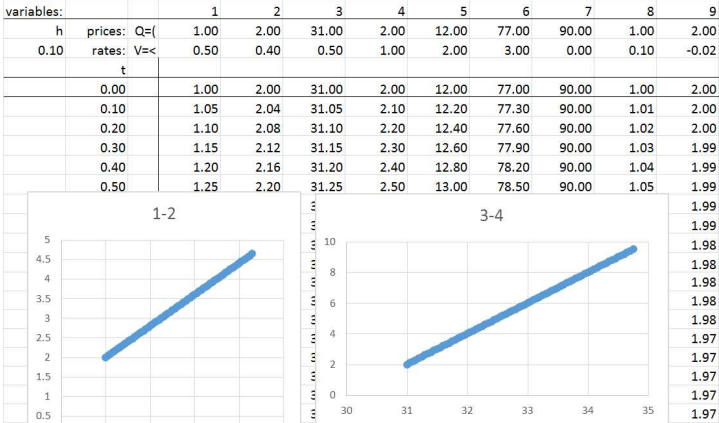
Stated for $m = 1$, the definition produces the familiar point-slope form! The rate of change is a single number (the slope) because the change is entirely within the y -axis. What has changed is the context: there are infinitely many directions in \mathbf{R}^2 for change. That is why the change and the rate of change is a vector. But the equation looks exactly the same... even though each letter may contain unlimited amount of information!

Example 4.1.3: prices

The definition applies to the abstract spaces. If \mathbf{R}^m is the space of prices (of stocks or commodities), we might have $m = 10,000$. The prices recorded continuously will produce a parametric curve and this curve might be a straight line. This happens when the prices are growing (or declining) *proportionally* but, possibly, at different rates. Also, in the short term this curve is likely to look like a straight line: the most recent change of each price is recorded is then the same change is predicted for the next time period. In each column we use the same recursive formula for the k th price:

$$x_k(t + \Delta t) = x_k(t) + v_k \Delta t,$$

where v_k is the k th rate of change.



The table is our 10,000-dimensional curve! Can we visualize such a curve in any way? We pick two columns at a time and plot that curve on the plane. Since these columns correspond to the axes, we are plotting a “shadow” of our curve cast on the corresponding coordinate plane. They are all straight lines.

Exercise 4.1.4

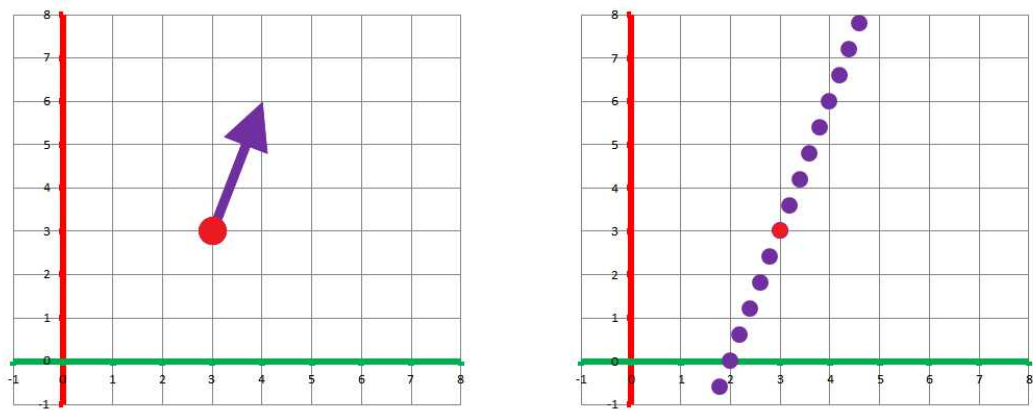
Find a parametric representation of the line through two distinct points P and Q .

In the physical space, a straight line is followed by an object when there are not forces at play. Even a constant force leads to acceleration which may change the direction of the motion.

Example 4.1.5: constant velocity

Recall these recursive formulas that gives the location as a function of time when the velocity is constant ($k = 0, 1, \dots$):

$$x : \quad p_{k+1} = p_k + v \Delta t$$
$$y : \quad q_{k+1} = q_k + u \Delta t$$



These quantities are now combined into points on the plane:

$$P_k = (p_k, q_k) ,$$

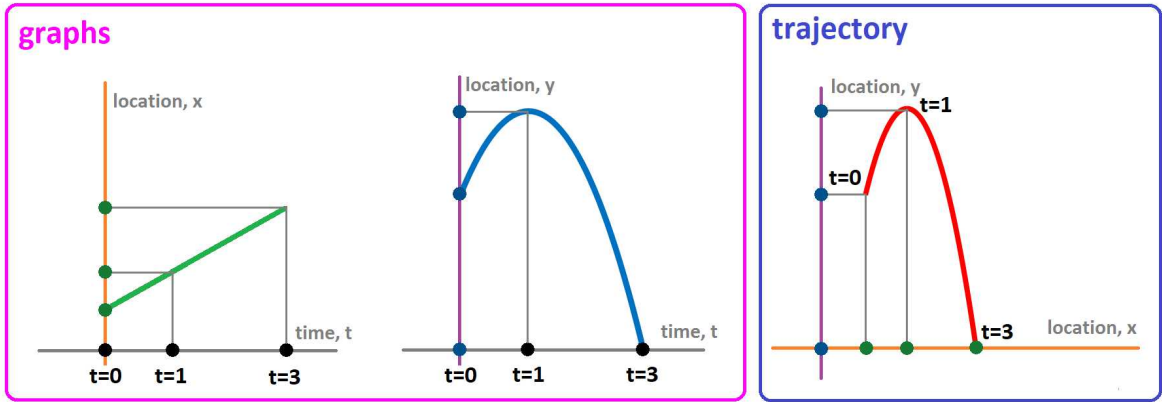
and the equations take a vector form too:

$$P_{k+1} = P_k + V_k \Delta t .$$

Example 4.1.6: thrown ball

Let’s review the dynamics of a thrown ball. A constant force causes the velocity to change linearly, just as the location in the last example. How does the location change this time?

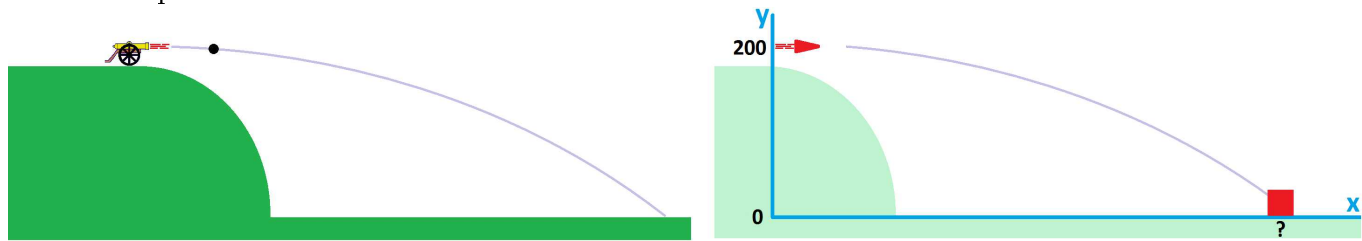
In the horizontal direction, as there is no force changing the velocity, the latter remains constant. Meanwhile, the vertical velocity is constantly changed by the gravity. The dependence of the height on the time is quadratic. The path of the ball will appear to an observer – from the right angle – as a curve:



A falling ball is subject to these accelerations, horizontal and vertical:

$$x : a_{k+1} = 0; \quad y : a_{k+1} = -g .$$

Now recall the setup considered previously: from a 200 feet elevation, a cannon is fired horizontally at 200 feet per second.



The initial conditions are:

- The initial location, $x : p_0 = 0$ and $y : p_0 = 200$.
- The initial velocity, $x : v_0 = 200$ and $y : v_0 = 0$.

Then we have two pairs of recursive equations – for the location in terms of the velocity and the velocity in terms of acceleration – independent of each other:

x

:

v_{k+1}

$= v_0$

p_{k+1}

$= p_k$

$+ v_k \Delta t$

y

:

u_{k+1}

$= v_k$

$- g \Delta t$

q_{k+1}

$= q_k$

$+ u_k \Delta t$

These are the formulas in the vector notation:

V_{k+1}

$= V_k$

$+ A$

$\cdot \Delta t$

P_{k+1}

$= P_k$

$+ V_{k+1}$

$\cdot \Delta t$

The advantage of the vector approach is that the choice of the coordinate system is no longer a concern!

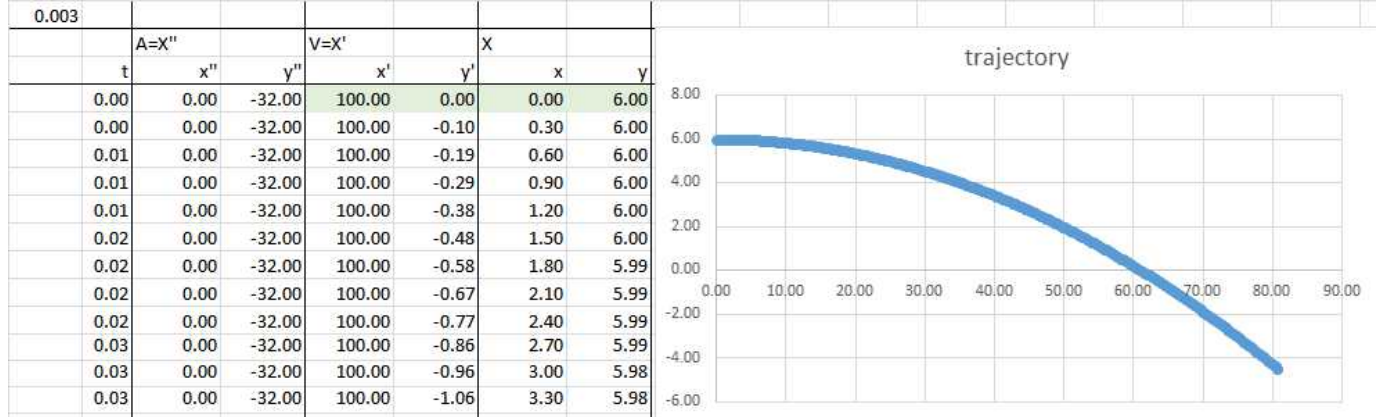
Example 4.1.7: recursive formulas

In dimension 2, for example, we don't have to align the x -axis with the direction of the throw and in dimension 3 we don't have to align the z -axis with the vertical direction.

Nonetheless, let's start with former case. A 6-foot man throws – straight forward – a ball with the speed of 100 feet per second. If the throw is along the x -axis and the y -axis is vertical, we have:

$$A = \langle 0, -32 \rangle, \quad V_0 = \langle 0, 100 \rangle, \quad P_0 = (6, 0).$$

This data goes into the first row of our table for the columns marked $x'', y'', x', y',$ and x, y respectively.



We apply the recursive formulas given above. In the spreadsheet,

- The velocity is computed from the velocity.
- The location is computed from the acceleration.

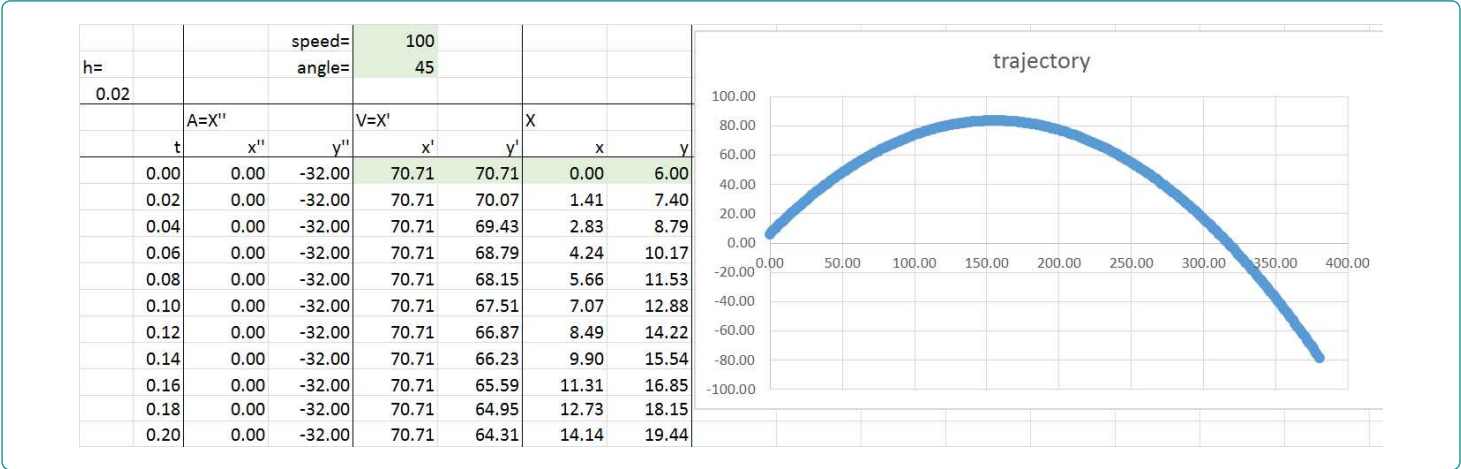
What is the difference of our *vectors* approach from our previous treatment of the flight of a ball? Instead of three columns for x'', x', x and then three columns for y'', y', y , one can see how the two components of acceleration, velocity, and location are combined into vectors contained in two columns each: x'', y'' , then x', y' , then x, y . The formula is almost the same as before:

=R[-1]C+(RC[-2]-R[-1]C[-2])*R2C1

Next an angled throw... The only change is the vector of initial velocity:

$$A = \langle 0, -32 \rangle, \quad V_0 = \langle 100 \cos \alpha, 100 \sin \alpha \rangle, \quad P_0 = (6, 0),$$

where α is the angle of the throw.



Example 4.1.8: continuous motion

Now the continuous case... Starting with the physics,

$$\begin{cases} x'' &= 0, \\ y'' &= -g, \end{cases}$$

we integrate – coordinate-wise – once:

$$\begin{cases} x' &= v_x, & x'(0) = v_x & \text{is the initial horizontal velocity,} \\ y' &= -gt + v_y, & y'(0) = v_y & \text{is the initial vertical velocity;} \end{cases}$$

and twice:

$$\begin{cases} x &= v_x t + p_x, & x(0) = p_x & \text{is the initial horizontal position,} \\ y &= -\frac{1}{2}gt^2 + v_y t + p_y, & y(0) = p_y & \text{is the initial vertical position.} \end{cases}$$

Thus, we have:

$$\begin{cases} \text{depth} &= \text{initial depth} + \text{initial horizontal velocity} \cdot \text{time} , \\ \text{height} &= \text{initial height} + \text{initial vertical velocity} \cdot \text{time} - \frac{1}{2}g \cdot \text{time}^2 . \end{cases}$$

We take this solution to the next level by assembling these components into vectors just as in the last example.

$$\text{location} = \text{initial location} + \text{initial velocity} \cdot \text{time} + \langle 0, -\frac{1}{2}g \cdot \text{time}^2 \rangle .$$

The last term needs work. The zero represents the zero horizontal acceleration while $-g$ is the vertical acceleration. Then the last term is the acceleration times $\frac{t^2}{2}$. Algebraically, we have:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} p_x \\ p_y \end{bmatrix} + \begin{bmatrix} v_x \\ v_y \end{bmatrix} \cdot t + \begin{bmatrix} 0 \\ -g \end{bmatrix} \cdot \frac{t^2}{2} .$$

The nature of the acceleration is irrelevant; we only need it to be constant.

Definition 4.1.9: parametric curve of uniformly accelerated motion

Suppose P_0 is a point in \mathbf{R}^m and V_0, A are vectors. Then the *parametric curve of uniformly accelerated motion through P_0 with the initial velocity of V and acceleration A* is:

$$P(t) = P_0 + V_0 \cdot t + A \cdot \frac{t^2}{2}$$

We have an extra term, that disappears when $A = 0$, in comparison to the uniform motion. Just as in the 1-dimensional case, a constant acceleration produces a quadratic motion!

Exercise 4.1.10

Show that the path of this parametric curve is a parabola.

The values of a function represented by a parametric curve lie in \mathbf{R}^m as *points* but can also be seen as *vectors*. For example, we can re-write the familiar parametric curve of points:

$$P(t) = P_0 + V_0 \cdot t + A \cdot \frac{t^2}{2},$$

as one of vectors:

$$R(t) = R_0 + V_0 \cdot t + A \cdot \frac{t^2}{2}.$$

Instead of passing through point P_0 it passes through the end point of vector $R_0 = OP_0$, which is the same thing. And, of course, the end of vector $R(t)$ is the point $P(t)$. The advantage of the latter approach is that it allows us to apply vector operations to the curves.

Example 4.1.11: circle transformed

Recall how we parametrized the unit circle using the angle as the parameter. Here, the x - and y -coordinates of a point at angle t is $\cos t$ and $\sin t$ respectively:

$$x = \cos t, \ y = \sin t.$$

The values of t may be the nodes of a partition of an interval such as $[0, 2\pi]$ or run through the whole interval.

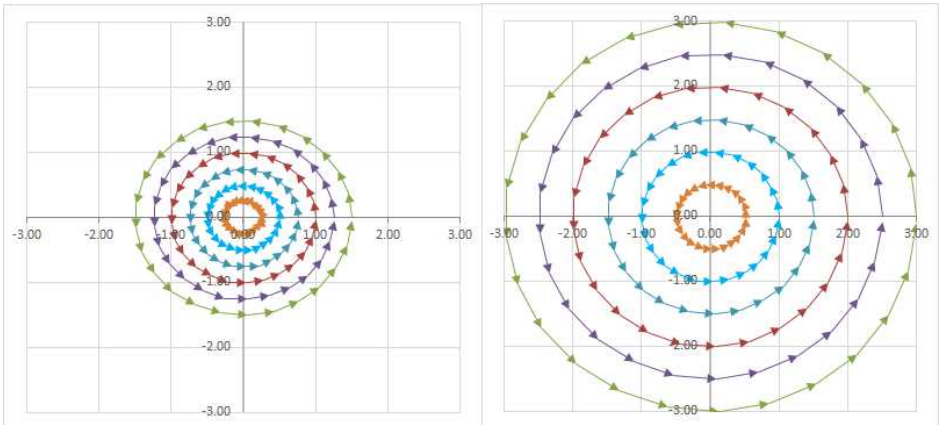
We can also look at this formula as a parametrization with respect to time. Then this is a record of motion with a constant speed or, in other words, a constant *angular* velocity. Now, this is the vector representation of this curve:

$$R(t) = \langle \cos t, \sin t \rangle.$$

So, applying vector operations to this curve will give as new curves, just as in the 1-dimensional case (Chapter 1PC-4). For example, using scalar multiplication by 2 on all vectors means *stretching radially* the whole space. We then discover that the curve in the plane given by:

$$Q(t) = 2R(t) = 2 \langle \cos t, \sin t \rangle,$$

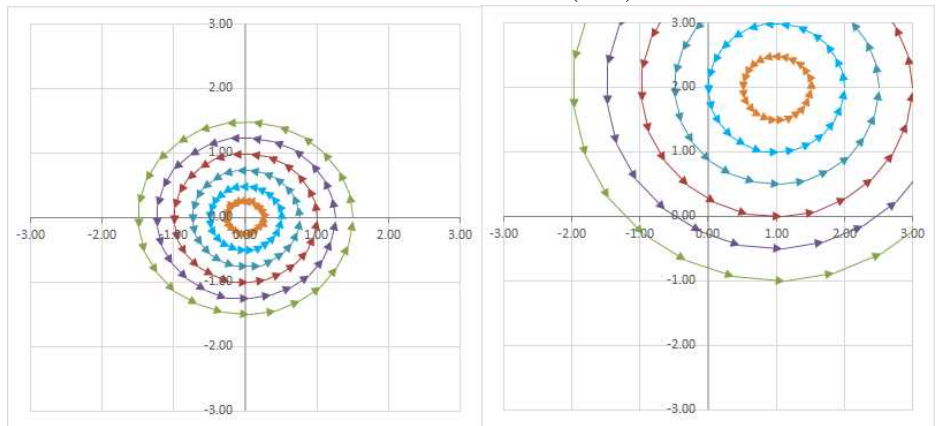
is a parametric curve of the circle of radius 2.



Similarly, using vector addition with $W = \langle 3, 1 \rangle$ on all vectors means *shifting* the whole space by this vector. We then discover that the curve in the plane given by:

$$S(t) = W + 2R(t) = (1, 2) + 2 \langle \cos t, \sin t \rangle ,$$

is a parametric curve of the circle of radius 2 centered at $(1, 2)$.



And so on with other *transformations* of the plane.

4.2. The predator-prey model

This is the IVP we have considered so far:

$$\frac{\Delta y}{\Delta t} = f(t, y), \quad y(t_0) = y_0 ,$$

and

$$y' = f(t, y), \quad y(t_0) = y_0 .$$

The equations show how the rate of change of y depends on t and y .

What if we have *two* variable quantities dependent on t ?

The simplest example is as follows:

- x is the horizontal location, and
- y is the vertical location.

We have already seen this simplest setting of free fall:

$$\left\{ \begin{array}{l} \frac{\Delta x}{\Delta t} = v_x, \\ \frac{\Delta y}{\Delta t} = v_y - gt, \end{array} \right. \quad \text{and} \quad \left\{ \begin{array}{l} x' = v_x, \\ y' = v_y - gt. \end{array} \right.$$

It is just as simple when arbitrary functions are in the right-hand sides of the equations (the continuous case is solved by integration). Here the rate of change of the location depends on the time t only.

More complex is the situation when the rate of change of the location depends on the location. When the former depends only on its own component of the location, the motion is described by this pair of ODEs:

$$\left\{ \begin{array}{l} \frac{\Delta x}{\Delta t} = g(x), \\ \frac{\Delta y}{\Delta t} = h(y), \end{array} \right. \quad \text{and} \quad \left\{ \begin{array}{l} x' = g(x), \\ y' = h(y). \end{array} \right.$$

The solution consists of two solutions to the two, unrelated, ODEs. We can then apply the methods of Chapter 3.

As an example of quantities that do interact, let’s consider the *predator-prey model*.

Let

- x be the number of rabbits and
- y be the number of foxes in the forest.

Let

- Δt be the fixed increment of time.

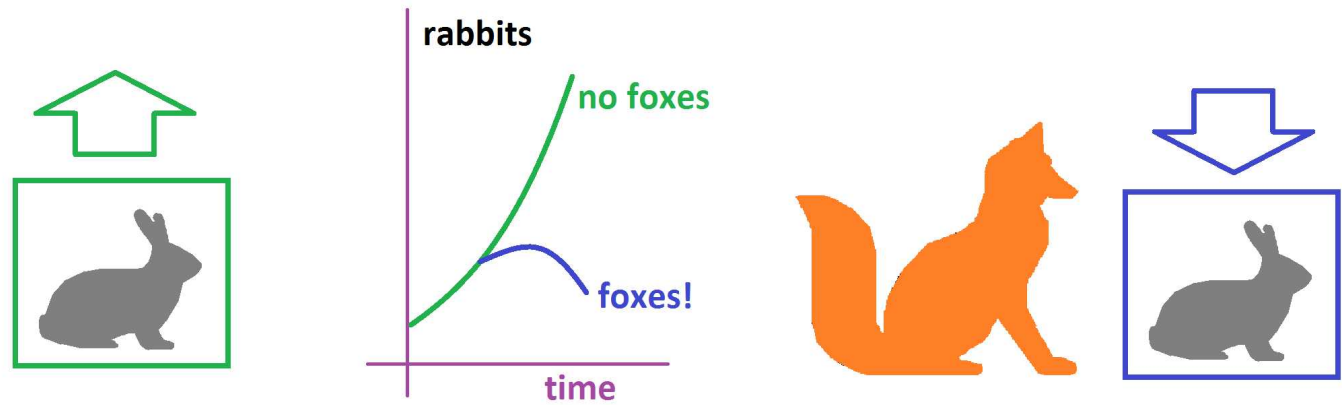
Even though time t is now discrete, the “number” of rabbits x or foxes y isn’t. Those are real numbers in our model. One can think of 0.1 rabbits as if the actual number is unknown but the likelihood that there is one somewhere is 10%.

We begin with the rabbits. There are two factors affecting their population.

First, we assume that they have an unlimited food supply and reproduce in a manner described previously – when there is no predator. In other words, the gain of the rabbit population per unit of time through their natural reproduction is proportional to the size of their current population. Therefore, we have:

$$\text{Rabbits' gain} = \alpha \cdot x \cdot \Delta t,$$

for some $\alpha > 0$.



Second, we assume that the rate of predation upon the rabbits to be proportional to the rate at which the rabbits and the foxes meet, which, in turn, is proportional to the sizes of their current populations, x and y . Therefore, we have:

$$\text{Rabbits' loss} = \beta \cdot x \cdot y \cdot \Delta t,$$

for some $\beta > 0$.

Combined, the change of the rabbit population over the period of time of length Δt is:

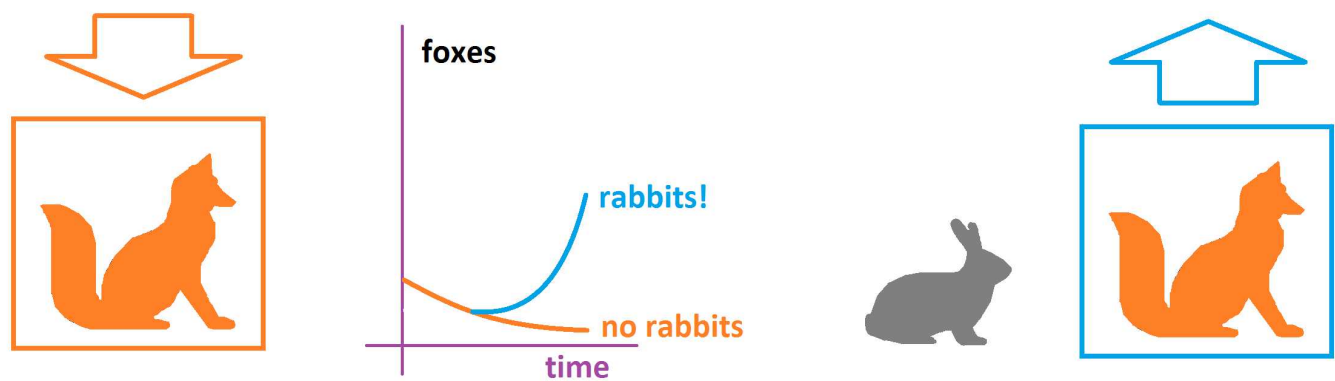
$$\Delta x = \alpha x \Delta t - \beta xy \Delta t.$$

We continue with the foxes. There are two factors affecting their population.

First, we assume that the foxes have only a limited food supply, i.e., the rabbits. The foxes die out geometrically in a manner described previously – when there is no prey. In other words, the loss of the fox population per unit of time through their natural death is proportional to the size of their current population. Therefore, we have:

$$\text{Foxes' loss} = \gamma \cdot y \cdot \Delta t,$$

for some $\gamma > 0$.



Second, we again assume that the rate of reproduction of the foxes is proportional to the rate of their predation upon the rabbits which is, as we know, proportional to the sizes of their current populations, x and y . Therefore, we have:

$$\text{Foxes' gain} = \delta \cdot x \cdot y \cdot \Delta t,$$

for some $\delta > 0$.

Combined, the change of the fox population over the period of time of length Δt is:

$$\Delta y = \delta xy \Delta t - \gamma y \Delta t.$$

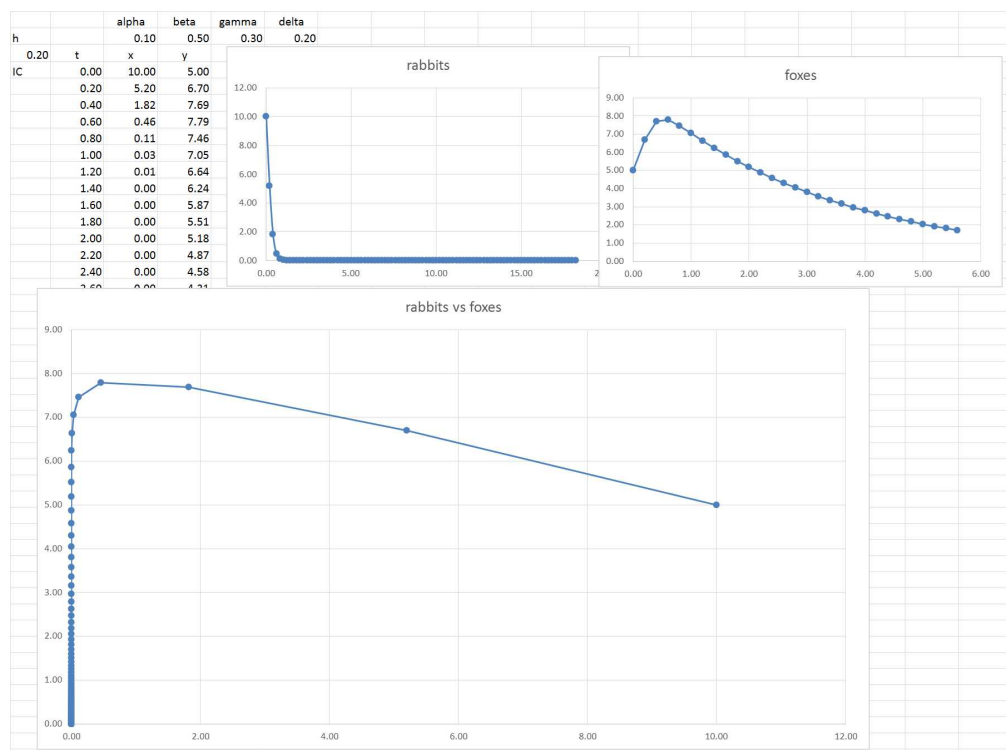
Putting these two together gives us a *discrete predator-prey model*:

$$\begin{cases} \Delta x = (\alpha x - \beta xy) \Delta t, \\ \Delta y = (\delta xy - \gamma y) \Delta t. \end{cases}$$

Then the spreadsheet formulas are for x and y respectively:

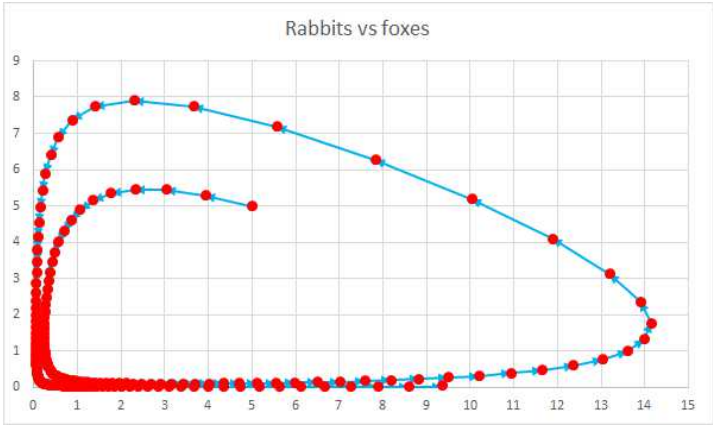
```
=R[-1]C+R2C3*R[-1]C*R3C1-R2C4*R[-1]C*R[-1]C[1]*R3C1  
=R[-1]C-R2C5*R[-1]C*R3C1+R2C6*R[-1]C*R[-1]C[-1]*R3C1
```

Let's take a look at an example of a possible dynamics ($\alpha = 0.10$, $\beta = 0.50$, $\gamma = 0.20$, $\delta = 0.20$, $h = 0.2$):



This is what we see. Initially, there are many rabbits and, with this much food, the number of foxes was growing: \uparrow . This was causing the number of rabbits to decline: \leftarrow . Combined, this is the direction of the system: \nwarrow . Later, the number of rabbits declines so much that, with so little food, the number of foxes also started to decline: \downarrow . At the end, both of the populations seem to have disappeared...

Another experiment shows that they can recover ($\alpha = 3, \beta = 2, \gamma = 3, \delta = 1, h = 0.03$):



In fact, we can see a repeating pattern.

Furthermore, with $\Delta t \rightarrow 0$, we have two, related, ODEs, a *continuous predator-prey model*:

$$\begin{cases} \frac{dx}{dt} = \alpha x - \beta xy, \\ \frac{dy}{dt} = \delta xy - \gamma y. \end{cases}$$

It approximates the discrete model. The equations are known as the *Lotka–Volterra equations*.

4.3. Qualitative analysis of the predator-prey model

To confirm our observations, we will carry out qualitative analysis. It is equally applicable to both the discrete model and the system of ODEs. Indeed, they both have the same right-hand side:

$$\begin{cases} \frac{\Delta x}{\Delta t} = \alpha x - \beta xy, \\ \frac{\Delta y}{\Delta t} = \delta xy - \gamma y, \end{cases} \quad \text{and} \quad \begin{cases} \frac{dx}{dt} = \alpha x - \beta xy, \\ \frac{dy}{dt} = \delta xy - \gamma y. \end{cases}$$

We will investigate the dynamics at all locations in the first quadrant of the tx -plane.

First, we find the locations where x has a zero derivative, i.e., $x' = 0$, which is the same as the locations where the discrete model leads to no change in x , i.e., $\Delta x = 0$. The condition is:

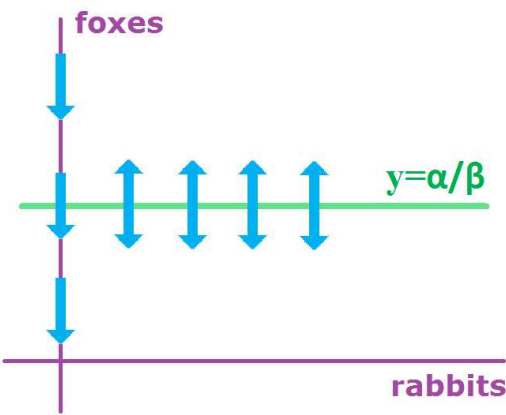
$$\alpha x - \beta xy = 0.$$

We solve the equation:

$$x = 0 \quad \text{or} \quad y = \alpha/\beta.$$

We discover that, first,

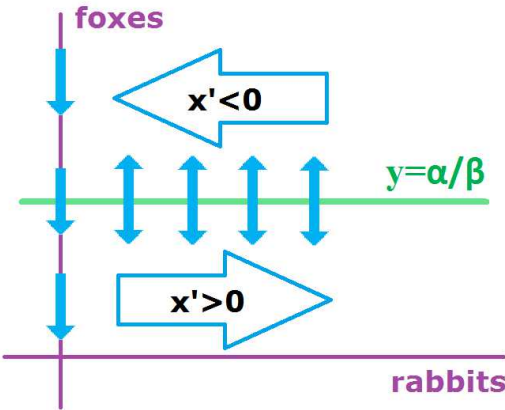
- $x = 0$ is a solution, and, second,
- the horizontal line $y = \alpha/\beta$ is crossed vertically by the solutions.



In other words, first, the foxes are dying out with no rabbits and, second, there may be a reversal in the dynamics of the rabbits at a certain number of foxes. To find out, solve the inequality:

$$x' > 0 \quad \text{or} \quad \Delta x > 0 \implies \alpha x - \beta xy > 0 \implies y < \alpha/\beta.$$

It follows that, indeed, the number of rabbits increases when the number of foxes is below α/β , otherwise it decreases.



Second, we find the locations where y has a zero derivative, i.e., $y' = 0$, which is the same as the locations where the discrete model leads to no change in y , i.e., $\Delta y = 0$. The condition is:

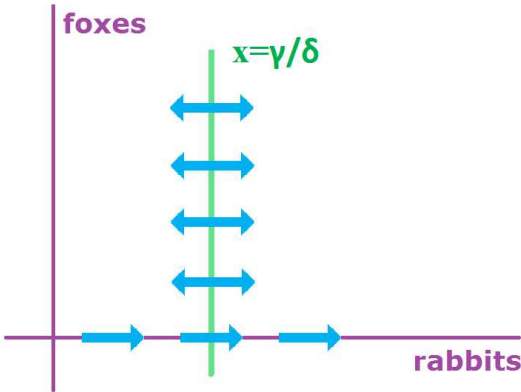
$$\delta xy - \gamma y = 0.$$

We solve the equation:

$$y = 0 \text{ or } x = \gamma/\delta.$$

We discover that, first,

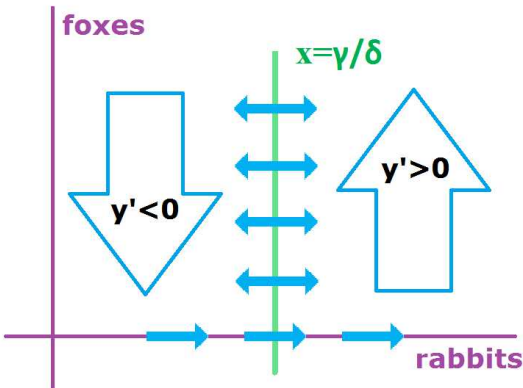
- $y = 0$ is a solution, and, second,
- the vertical line $x = \gamma/\delta$ is crossed horizontally by the solutions.



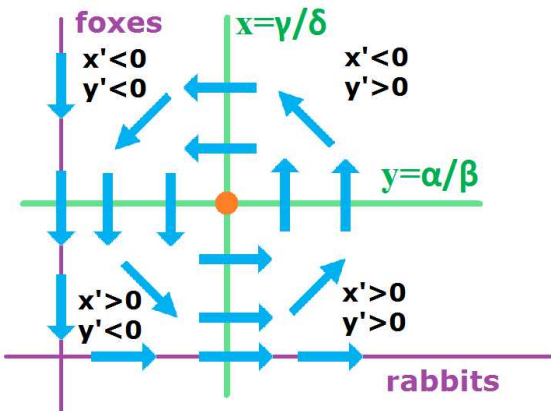
In other words, first, the rabbits thrive with no foxes and, second, there may be a reversal in the dynamics of the foxes at a certain number of rabbits. To find out, solve the inequality:

$$y' > 0 \text{ or } \Delta y > 0 \implies \delta xy - \gamma y > 0 \implies x > \gamma/\delta.$$

It follows that, indeed, the number of foxes increases when the number of rabbits is above γ/δ , otherwise it decreases.

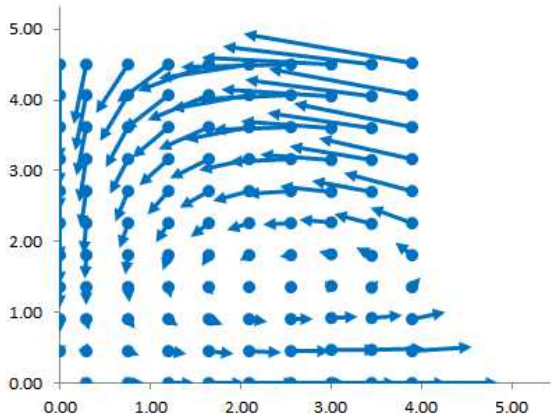


Now we put these two parts together. We have four sectors in the first quadrant determined by the four different choices of the signs of x' and y' , or Δx and Δy :



In either case, the result is a rough description of the dynamics on the local level: if this is the current state, then this is the direction it is going. It is a *vector field*!

We visualize this vector field with the same parameters as before:



The arrows aren't meant yet to be connected into curves to produce solutions. The only four distinct solutions we know for sure are the following:

- the decline of the foxes in the absence rabbits – on the y -axis;
- the explosion of the rabbits in the absence of foxes – on the x -axis;
- the freezing of both rabbits and foxes at the special levels – in the middle; and also
- the freezing of both rabbits and foxes at the zero level.

Either of the last two is a constant solution called an *equilibrium*. The main, non-zero, equilibrium is:

$$x(t) = \gamma/\delta, \; y(t) = \alpha/\beta.$$

What about the rest of the solutions?

In order to confirm that the solutions circle the main equilibrium, we need a more precise analysis. In each of the four sectors, the monotonicity of the solutions has been proven. However, it is still possible that some solutions will stay within the sector when one or both of x and y behave asymptotically:

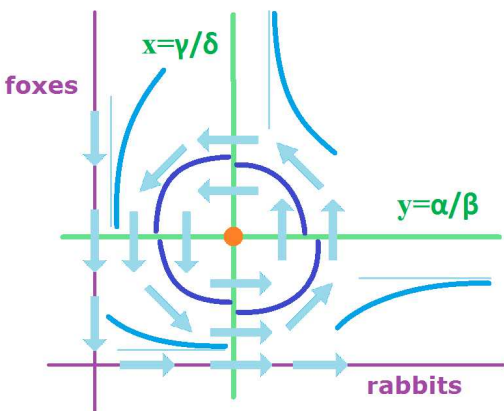
$$x(t) \rightarrow a \text{ and/or } y(t) \rightarrow b \text{ as } t \rightarrow \infty.$$

Since both functions are monotonic, this implies that

$$x'(t) \rightarrow 0 \text{ and/or } y'(t) \rightarrow 0 \text{ as } t \rightarrow \infty,$$

and the same for Δx and Δy . We can show that this is impossible. For example, suppose we start in the bottom right sector, i.e., the initial conditions are:

- $x(0) = p > \gamma/\delta$;
- $y(0) = q < \alpha/\beta$.



Then, for as long as the solution is in this sector, we have

- $x' > 0 \implies x > p$
- $y' > 0 \implies y > q$

Therefore,

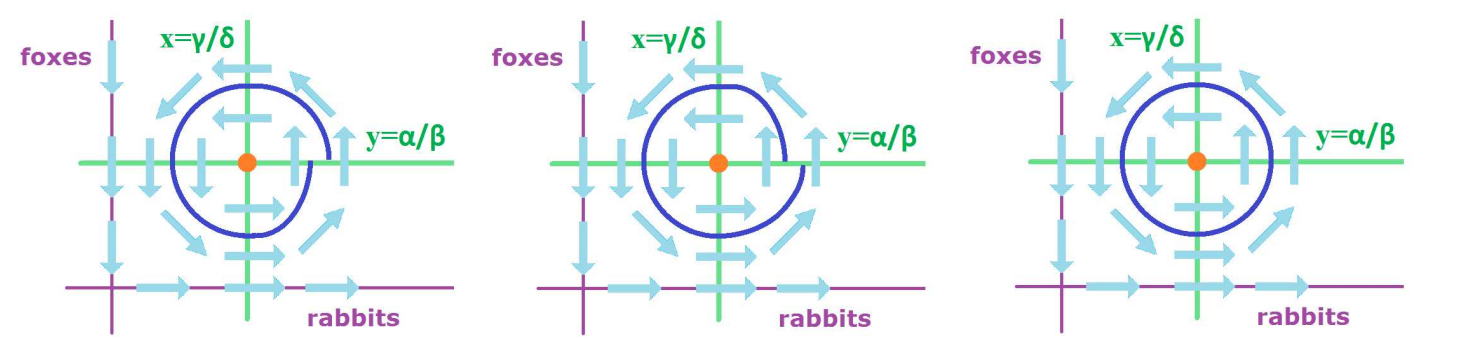
$$y' = y(\delta x - \gamma) > q(\delta p - \gamma) > 0.$$

This number is a gap between y' and 0. Therefore, y' cannot diminish to 0, and the same is true for Δy . It follow that the solution will reach the line $y = \alpha/\beta$ “in finite time”.

Exercise 4.3.1

Prove the analogous facts about the three remaining sectors.

We have demonstrated that a solution will go around the main equilibrium, but when it comes back, will it be closer to the center, farther, or will it come to the same location?



The first option is indicated by our spreadsheet result. Next, we set the discrete model aside and concentrate on solving analytically our system of ODEs to answer the question, is this *a cycle or a spiral*?

4.4. Solving the Lotka–Volterra equations

We would like to *eliminate time* from the equations ($x > 0, y > 0$):

$$\begin{cases} \frac{dx}{dt} = \alpha x - \beta xy, \\ \frac{dy}{dt} = \delta xy - \gamma y. \end{cases}$$

This step is made possible by the following *trick*. We interpret these derivatives in terms of the differential forms:

$$\begin{aligned} dx &= (\alpha x - \beta xy)dt \implies dt = \frac{dx}{\alpha x - \beta xy} \\ dy &= (\delta xy - \gamma y)dt \implies dt = \frac{dy}{\delta xy - \gamma y} \end{aligned}$$

Therefore,

$$dt = \frac{dx}{\alpha x - \beta xy} = \frac{dy}{\delta xy - \gamma y}.$$

We next *separate variables*:

$$\frac{\delta x - \gamma}{x}dx = \frac{\alpha - \beta y}{y}dy.$$

We integrate:

$$\int \left(\delta - \frac{\gamma}{x} \right) dx = \int \left(\frac{\alpha}{y} - \beta \right) dy,$$

and we have:

$$\delta x - \gamma \ln x = \alpha \ln y - \beta y + C.$$

The system is *solved*!

But what does this equation mean?

Every solution $x = x(t)$ and $y = y(t)$, when substituted into the function

$$G(x, y) = \delta x - \gamma \ln x - \alpha \ln y + \beta y,$$

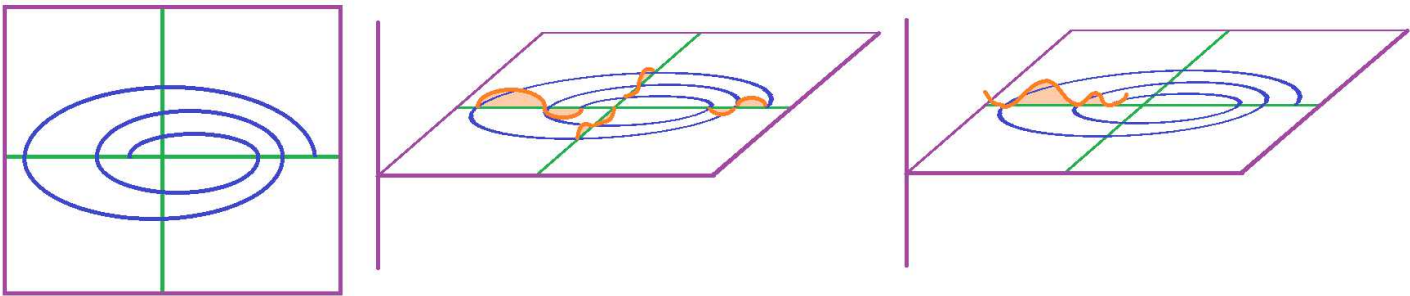
produces a constant. In other words, this parametric curve is a *level curve* of $z = G(x, y)$.

Exercise 4.4.1

Prove the last statement.

Once we have no derivatives left, we declare the system solved even though only implicitly. Even though we don't have explicit formulas for $x = x(t)$ or $y = y(t)$, we can use what we have to further study the qualitative behavior of the system.

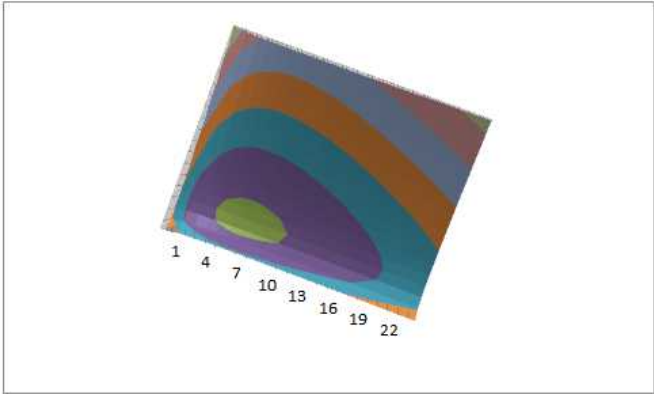
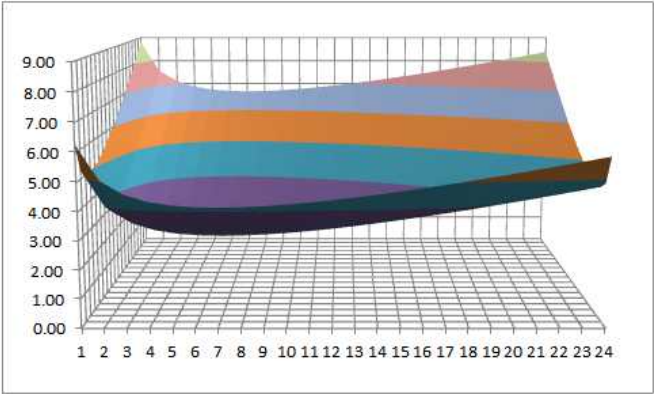
For example, the fact that this is a level curve already suggests that the parametric curve is *not a spiral*:



Just try to imagine such a surface that its level curves are spirals.

We turn instead to the actual function.

First, we plot it with a spreadsheet ($\alpha = 3, \beta = 2, \gamma = 3, \delta = 1$):



The level curves are visible. Some of them are clearly circular and others aren't. The reason is that they aren't shown all the way to the axes because the value of G rises so quickly (in fact, asymptotically).

As expected, the surface seems to have a single minimum point. Let's prove that algebraically:

$$\begin{aligned} \frac{\partial G}{\partial x}(x, y) &= \delta - \gamma/x, \\ \frac{\partial G}{\partial y}(x, y) &= -\alpha/y + \beta. \end{aligned}$$

We next find the extreme points of G . We set the two derivatives equal to zero and solve for x and y :

$$\begin{aligned}\frac{\partial G}{\partial x}(x,y) &= \delta - \gamma/x = 0 \implies x = \frac{\gamma}{\delta} \\ \frac{\partial G}{\partial y}(x,y) &= -\alpha/y + \beta = 0 \implies y = \frac{\alpha}{\beta}\end{aligned}$$

This point is indeed our main *equilibrium point*. The surface here has a horizontal tangent plane. We have also demonstrated that there are no others points like that!

But could this be a maximum point? Just as $y = x^3$ crosses the x -axis at 0 degrees, a surface can cross its tangent plane.

We compute the second derivatives:

$$\begin{aligned}\frac{\partial^2 G}{\partial x^2}(x,y) &= \gamma \frac{1}{x^2} > 0, \\ \frac{\partial^2 G}{\partial y^2}(x,y) &= \alpha \frac{1}{y^2} > 0.\end{aligned}$$

Both are positive throughout, therefore, either of the cross-sections of the surface along the axes have a minimum point here and it has to stay on one side of the plane. However, it might cross it at other directions. For example, this still might be a saddle point! We invoke the *Second Derivative Test* from [Chapter 2DC-5](#) to resolve this.

We consider the *Hessian matrix* (discussed in [Chapter 4HD-3](#)) of G . It is the 2×2 matrix of the four partial derivatives of G :

$$H(x,y) = \begin{pmatrix} \frac{\partial^2 G}{\partial x^2} & \frac{\partial^2 G}{\partial x \partial y} \\ \frac{\partial^2 G}{\partial y \partial x} & \frac{\partial^2 G}{\partial y^2} \end{pmatrix} = \begin{pmatrix} \gamma \frac{1}{x^2} & 0 \\ 0 & \alpha \frac{1}{y^2} \end{pmatrix}.$$

Here, in addition, we have the mixed second derivatives:

$$\frac{\partial^2 G}{\partial x \partial y}(x,y) = \frac{\partial^2 G}{\partial y \partial x}(x,y) = 0.$$

Next, we look at the determinant of the Hessian:

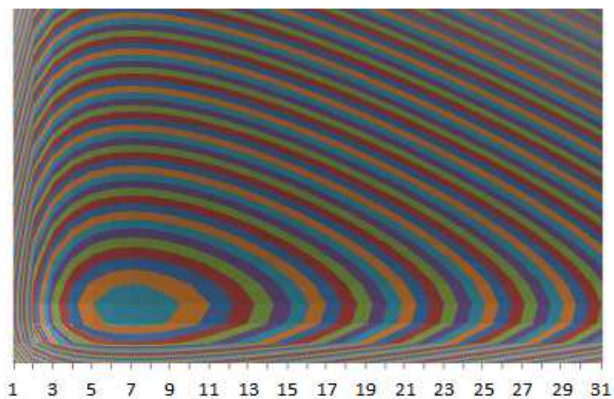
$$D(x,y) = \det(H(x,y)) = \left(\gamma \frac{1}{x^2}\right) \cdot \left(\alpha \frac{1}{y^2}\right) = \frac{\alpha \gamma}{x^2 y^2} > 0.$$

It's positive! Therefore, the point is a minimum.

We conclude that every level curve of G , i.e., a solution of the system, goes around the equilibrium and comes back to create a cycle.

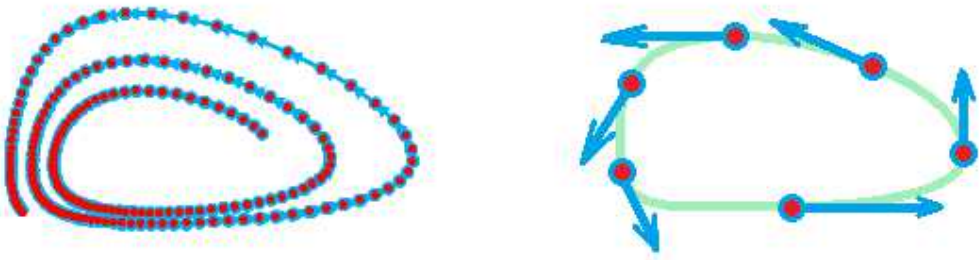
An easier, but more *ad hoc*, way to reach this conclusion is to imagine that a solution starts on, say, the line $y = \alpha/\beta$ at $x = x_0$ to the right of the equilibrium and then comes back to the line at $x = x_1$. Since this is a level curve, we have $G(x_0, \alpha/\beta) = G(x_1, \alpha/\beta)$. According to *Rolle's Theorem* from [Chapter 2DC-5](#), this contradicts our conclusion that $\frac{\partial G}{\partial x} > 0$ along this line.

Plotted with the same parameters, this is what these curves look like:



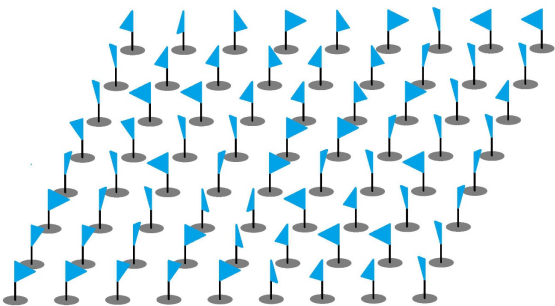
Exercise 4.4.2

Prove that the predator-prey discrete model, i.e., Euler's method for this system of ODEs, produces solutions that spiral *away* from the equilibrium. Hint:

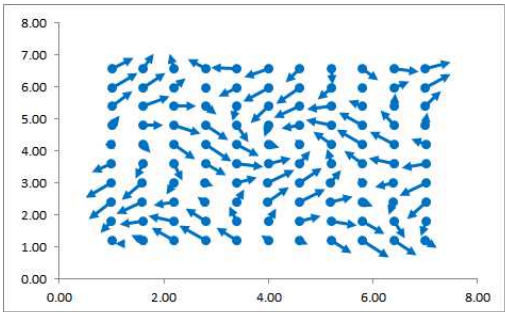


4.5. Vector fields and systems of ODEs

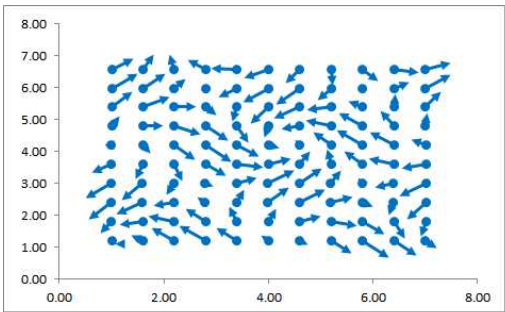
Numerous processes are modeled by systems of ODEs. For example, if little flags are placed on the lawn, then their directions taken together represent a system of ODEs, while the (invisible) air flow is the solutions of this system.



A similar idea is used to model a fluid flow. The dynamics of each particle is governed by the velocity of the flow, at each location, the same at every moment of time.



To *solve* such a system would require tracing the path of every particle of the liquid.
Let's review the discrete model of a flow: given a flow on a plane, trace a single particle of this stream.



For both coordinates, x and y , the following table is being built. The initial time t_0 and the initial location p_0 are placed in the first row of the spreadsheet. As we progress in time and space, new numbers are placed in the next row of our spreadsheet:

$$t_n, \, v_n, \, p_n, \, n = 1, 2, 3, \dots$$

The following *recursive formulas* are used:

$$t_{n+1} = t_n + \Delta t$$

and

$$p_{n+1} = p_n + v_{n+1} \cdot \Delta t.$$

The result is a growing table of values:

	iteration n	time t_n	velocity v_n	location p_n
initial:	0	3.5	--	22
	1	3.6	33	25.3

	1000	103.5	4	336

So, instead of two (velocity – location, as before), there will be four main columns when the motion is two-dimensional and six when it is three-dimensional:

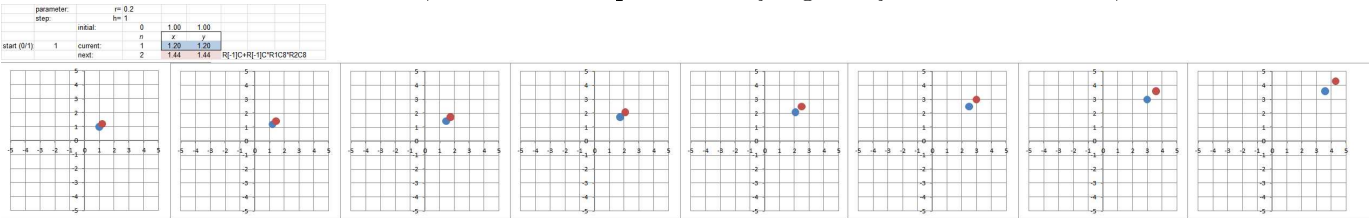
	time	horiz.	horiz.	vert.	vert.	...
n	t	vel. x'	loc. x	vel. y'	loc. y	...
0	3.5	--	22	--	3	...
1	3.6	33	25.3	4	3.5	...
...
1000	103.5	4	336	66	4	...
...

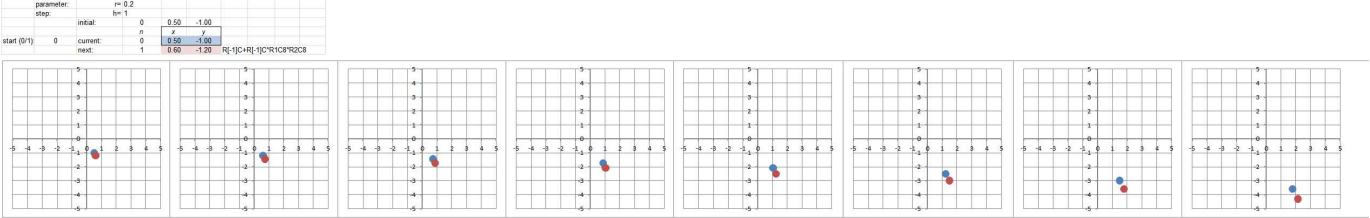
Example 4.5.1: simulation

Recall the examples such flows. If the velocity of the flow is proportional to the location:

$$v_{n+1} = 0.2 \cdot p_n,$$

for both horizontal and vertical, the result is particles flying away from the center, faster and faster:





If the horizontal velocity is proportional to the vertical location and the vertical velocity proportional to the negative of the horizontal location, the result resembles rotation:

Definition 4.5.2: solution of system of ODEs

A *solution* of a system of ODEs is a pair of functions $x = x(t)$ and $y = y(t)$ (a parametric curve) with either one differentiable on an open interval I such that for every t in I we have:

$$\begin{cases} x'(t) &= f(x(t), y(t)) \\ y'(t) &= g(x(t), y(t)) \end{cases}$$

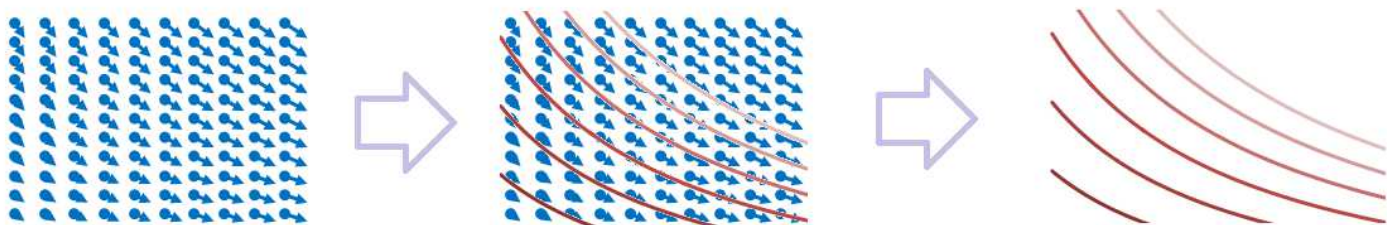
or abbreviated:

$$\begin{cases} x' &= f(x, y) \\ y' &= g(x, y) \end{cases}$$

The vector field of this system is the slope field of the ODE.
How do we visualize the solutions of such a system? With a single ODE,

$$y' = f(t) \implies y = y(t),$$

we simply plot their *graphs*, i.e., the collections of points $(t, y(t))$, of some of them on the ty -plane. This time, the solutions are parametric curves! Their graphs, i.e., the collections of $(t, x(t), y(t))$ lie in the 3-dimensional txy -space. That is why, we, instead, plot their *images*, i.e., the collections of points $(x(t), y(t))$ on the xy -plane. In the theory of ODEs, they are also known as *trajectories*, or paths.



Then the vectors of the vector field are tangent to these trajectories.
Since the vector field is independent of t , such a representation is often sufficient as explained by the following result.

Theorem 4.5.3: Shifted Solutions

If $x = x(t)$, $y = y(t)$ is a solution of the system of ODEs, then so is $x = x(t + s)$, $y = y(t + s)$ for any real s .

It is also important to be aware of the fact that the theory of systems of ODEs “includes” the theory of single ODEs. Recall, first, how the graph of every function can be represented by the trajectory of a parametric curve:

$$y = r(x) \longrightarrow \begin{cases} x = t, \\ y = r(x). \end{cases}$$

Similarly, the solutions of every time-independent ODE can be represented by the trajectories of a system of two ODEs:

$$y' = g(x) \longrightarrow \begin{cases} x' &= 1, \\ y' &= g(y). \end{cases}$$

Definition 4.5.4: initial value problem

For a given system of ODEs and a given triple (t_0, x_0, y_0) , the *initial value problem*, or an IVP, is

$$\begin{cases} x' = f(x, y), \\ y' = g(x, y); \end{cases} \quad \begin{cases} x(t_0) = x_0, \\ y(t_0) = y_0; \end{cases}$$

and its *solution* is a solution of the ODE that satisfies the *initial condition* above.

By the last theorem, the value of t_0 doesn't matter; it can always be chosen to be 0.

Definition 4.5.5: existence property

We say that a system of ODEs satisfies the *existence* property at a point (t_0, x_0, y_0) when the IVP above has a solution.

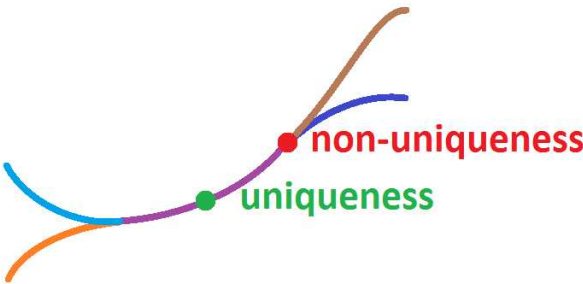
If your model of a real-life process doesn't satisfy existence, it reflects limitations of your model. It is as if the process starts but never continues.

Definition 4.5.6: uniqueness property

We say that an ODE satisfies the *uniqueness* property at a point (t_0, x_0, y_0) if every pair of solutions, (x_1, y_1) , (x_2, y_2) , of the IVP above are equal,

$$x_1(t) = x_2(t), \quad y_1(t) = y_2(t) ,$$

for every t in some open interval that contains t_0 .



If your model of a real-life process doesn't satisfy uniqueness, it reflects limitations of your model. It's as if you have all the data but can't predict even the nearest future.

Thus systems of ODEs produce families of curves as the sets of their solutions. Conversely, if a family of curves is given by an equation with a single parameter, we may be able to find a system of ODEs for it.

Example 4.5.7: anti-differentiation

The family of vertically shifted graphs,

$$y = x^2 + C ,$$

creates an ODE if we differentiate (implicitly) with respect to t :

$$y' = 2xx' .$$

Since these are just functions of x , we can choose $x = t$. This is a possible vector field for this family:

$$\begin{cases} x' = 1, \\ y' = 2x. \end{cases}$$

Example 4.5.8: exponential case

The family of stretched exponential graphs,

$$y = Ce^x,$$

creates an ODEs:

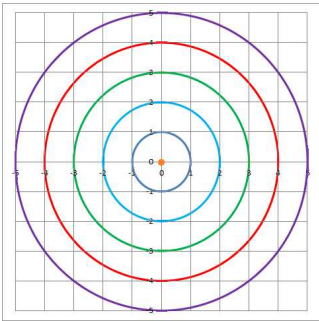
$$y' = Ce^x x'.$$

This is a possible vector field for this family:

$$\begin{cases} x' &= 1, \\ y' &= Ce^x. \end{cases}$$

Example 4.5.9: family of circles

What about these concentric circles?



(In the case when $C = 0$, we have the origin.) They are given by

$$x^2 + y^2 = C \geq 0.$$

We differentiate (implicitly) with respect to t :

$$2xx' + 2yy' = 0.$$

We choose what x' , y' might be equal to in order for the two terms to cancel. This is a possible vector field for this family:

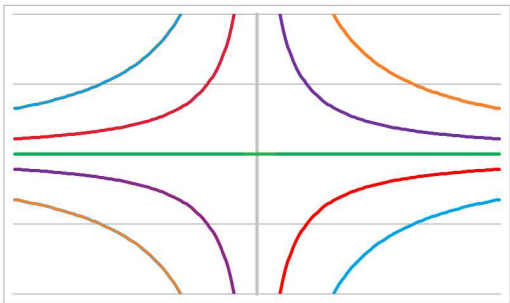
$$\begin{cases} x' &= -y, \\ y' &= x. \end{cases}$$

Example 4.5.10: family of hyperbolas

These hyperbolas are given by these equations:

$$xy = C.$$

(In the case when $C = 0$, we have the two axes.)



Then, we have an ODE:

$$x'y + xy' = 0.$$

This is a possible vector field for this family:

$$\begin{cases} x' &= x, \\ y' &= -y. \end{cases}$$

None of the examples have problems with either existence or uniqueness – in contrast to the corresponding ODEs.

The proofs of the following important theorems lie outside the scope of this book.

Theorem 4.5.11: Existence

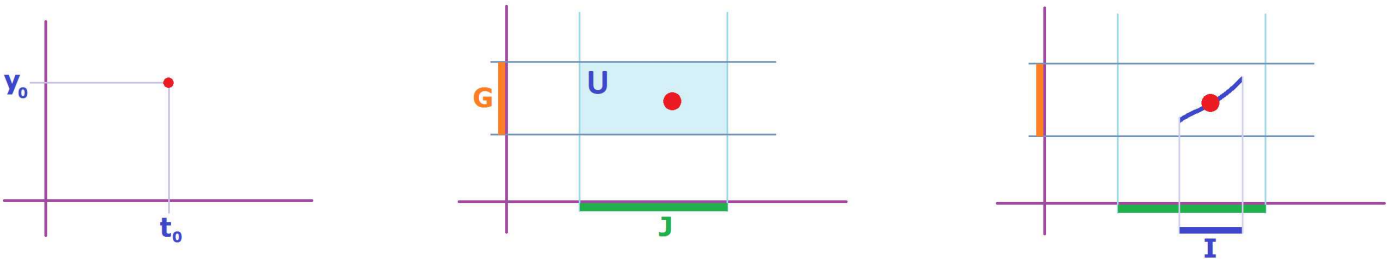
Suppose (x_0, y_0) is a point on the xy -plane and suppose:

- H is an open interval that contains x_0 .
- G is an open interval that contains y_0 .

Suppose also that functions $z = f(x, y)$ and $z = g(x, y)$ of two variables are continuous with respect to x and y on $H \times G$. Then the system of ODE,

$$\begin{cases} x' &= f(x, y), \\ y' &= g(x, y). \end{cases}$$

satisfies the existence property at (t_0, x_0, y_0) for any t_0 .



Theorem 4.5.12: Uniqueness

Suppose (x_0, y_0) is a point on the xy -plane and suppose:

- H is an open interval that contains x_0 .
- G is an open interval that contains y_0 .

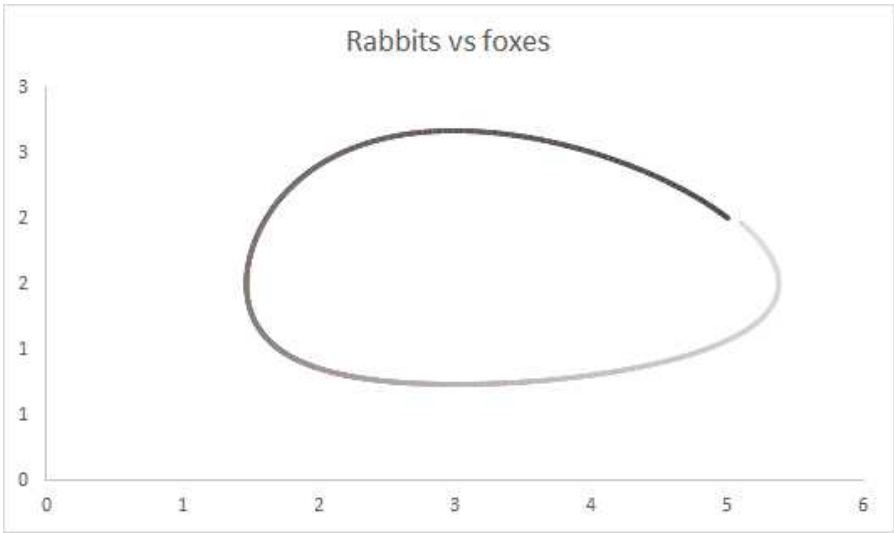
Suppose also that function $z = f(x, y)$ and $z = g(x, y)$ of two variables are differentiable with respect to x and y on $H \times G$. Then the system of ODEs,

$$\begin{cases} x' &= f(x, y), \\ y' &= g(x, y). \end{cases}$$

satisfies the uniqueness property at (t_0, x_0, y_0) for any t_0 .

4.6. Discrete systems of ODEs

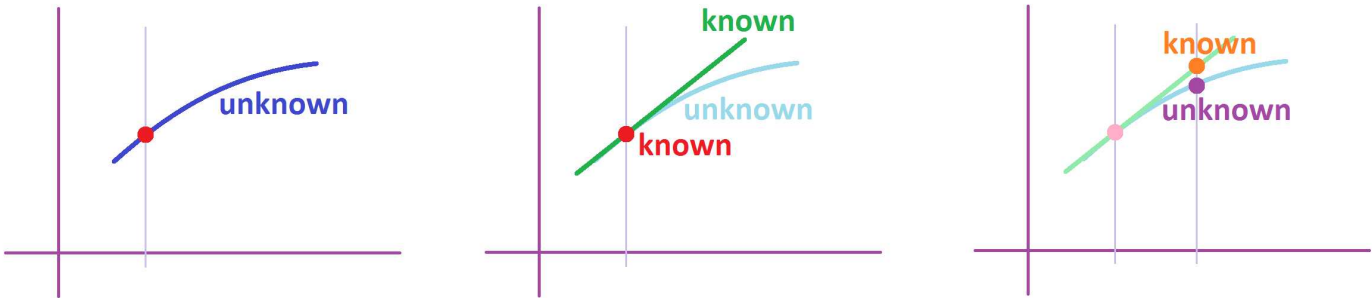
Discrete ODEs approximate and are approximated by continuous ODEs. The same is true for systems. For example, the discrete system for the predator-prey model produces this almost exactly cyclic path:



In other words, Euler’s method is capable of tracing solutions very close the ones of the ODE it came from. Just as in the 1-dimensional case, the IVP tells us:

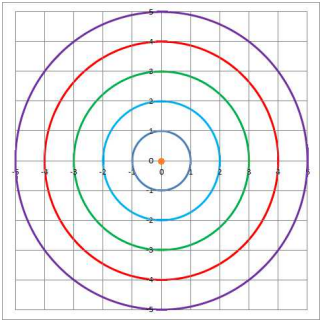
- where we are (the initial condition), and
- the direction we are going (the ODE).

Just as before, the unknown solution is replaced with its *best linear approximation*.



Example 4.6.1: family of circles

Let’s consider again these concentric circles:



They are the solutions of the ODEs:

$$\begin{cases} x' = y, \\ y' = -x. \end{cases}$$

We choose the increment of t :

$$\Delta t = 1.$$

We start with this initial condition:

$$t_0 = 0, \quad x_0 = 0, \quad y_0 = 2.$$

We substitute these two numbers into the equations:

$$\begin{cases} x' = 2, \\ y' = 0; \end{cases}$$

This is the direction we will follow. The increments are

$$\begin{cases} \Delta x = 2 \cdot \Delta t = 2 \cdot 1 = 2, \\ \Delta y = 0 \cdot \Delta t = 0 \cdot 1 = 0. \end{cases}$$

Our next location on the xy -plane is then:

$$\begin{cases} x_1 = x_0 + \Delta x = 0 + 2 = 2, \\ y_1 = y_0 + \Delta y = 2 + 0 = 2. \end{cases}$$

A new initial condition appears:

$$x_0 = 2, \quad y_0 = 2.$$

We again substitute these two numbers into the equations:

$$\begin{cases} x' &= 2, \\ y' &= -2. \end{cases}$$

producing the direction we will follow. The increments are

$$\begin{cases} \Delta x = 2 \cdot \Delta t = 2 \cdot 1 = 2, \\ \Delta y = -2 \cdot \Delta t = -2 \cdot 1 = -2. \end{cases}$$

Our next location on the xy -plane is then:

$$\begin{cases} x_2 = x_1 + \Delta x = 2 + 2 = 4, \\ y_2 = y_1 + \Delta y = 2 + (-2) = 0. \end{cases}$$

One more IVP:

$$x_2 = 0, \quad y_2 = -4.$$

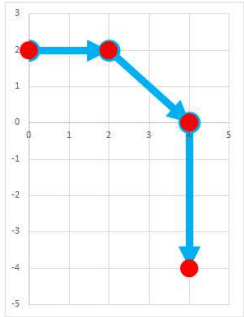
The increments are

$$\begin{cases} \Delta x = 0 \cdot \Delta t = 0 \cdot 1 = 0, \\ \Delta y = -4 \cdot \Delta t = -4 \cdot 1 = -4. \end{cases}$$

Our next location on the xy -plane is then:

$$\begin{cases} x_3 = x_2 + \Delta x = 4 + 0 = 4, \\ y_3 = y_2 + \Delta y = 0 - 4 = -4. \end{cases}$$

These four points form a very crude approximation of one of our circular solutions:



They are clearly spiraling away from the origin.

In terms of motion, this is our plan:

- At our current location and current time, we examine the ODE to find the velocity and then move accordingly to the next location.

Definition 4.6.2: Euler solution

The *Euler solution* with increment $\Delta t > 0$ of the IVP

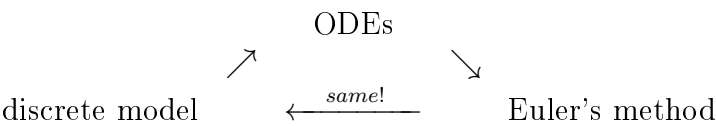
$$\begin{cases} x' &= f(x,y) \\ y' &= g(x,y) \end{cases} \quad \begin{cases} x(t_0) &= x_0 \\ y(t_0) &= y_0 \end{cases}$$

is the two sequences $\{x_n\}$ and $\{y_n\}$ of real numbers given by:

$$\begin{cases} x_{n+1} &= x_n + f(x_n, y_n) \cdot \Delta t \\ y_{n+1} &= y_n + g(x_n, y_n) \cdot \Delta t \end{cases}$$

where $t_{n+1} = t_n + \Delta t$.

Once again, if we derived our ODEs from a discrete model (via $\Delta t \rightarrow 0$), Euler’s method will bring us right back to it:



Example 4.6.3: spreadsheet

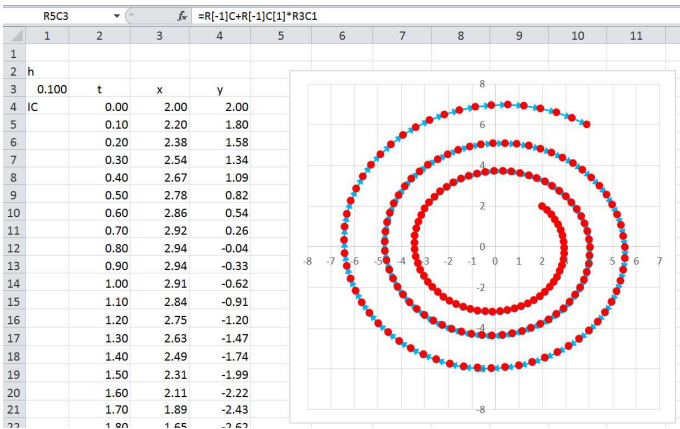
Let’s now carry out this procedure with a spreadsheet. The formulas for x_n and y_n are respectively:

=R[-1]C+R[-1]C[1]*R3C1

and

=R[-1]C-R[-1]C[-1]*R3C1

These are the results:



In contrast to the case of a single ODE, the approximations do not behave erratically close to the x -axis. The reason is that there is no division by y anymore.

Example 4.6.4: family of hyperbolas

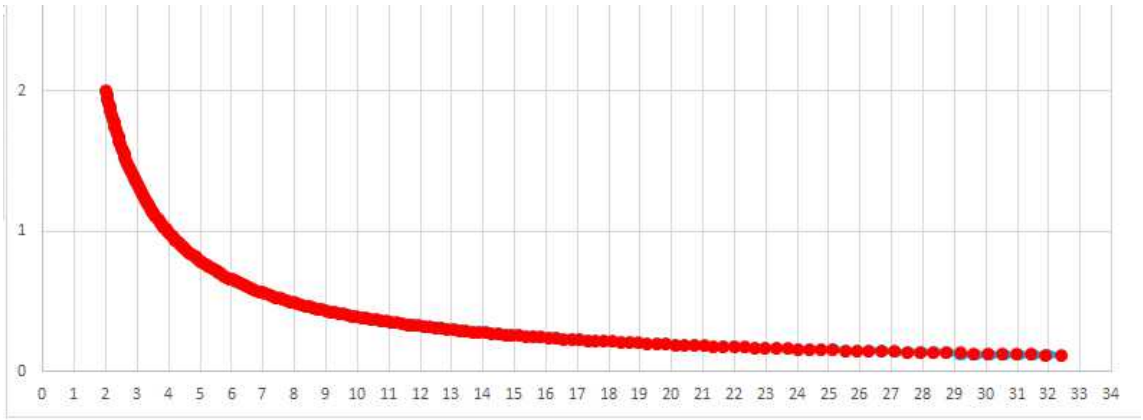
Let’s consider again these hyperbolas:

$$xy = C.$$

They are the solutions of the system:

$$\begin{cases} x' &= x, \\ y' &= -y. \end{cases}$$

An Euler solution is shown below:



However, is this asymptotic convergence toward the x -axis or do they merge?

4.7. Qualitative analysis of systems of ODEs

Since Euler’s method depends on the value of h , the increment of t , then, even with smaller and smaller values of h , the result remains a mere approximation. Meanwhile, *qualitative* analysis collects information about the solutions without solving the system – either analytically or numerically. The result is fully accurate but very broad descriptions of the solutions.

Example 4.7.1: 1d qualitative analysis

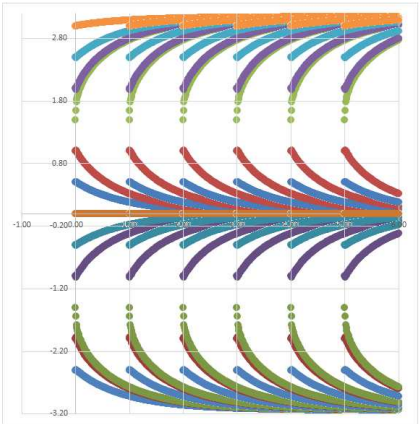
Let’s review an example of a single ODE from last section:

$$y' = -\tan(y).$$

This is what we conclude about the strip $[-\pi/2, \pi/2]$:

- For $-\pi/2 < y < 0$, we have $y' = -\tan y > 0$ and, therefore, $y \nearrow$.
- For $y = 0$, we have $y' = -\tan y = 0$ and, therefore, y is a constant solution.
- For $0 < y < \pi/2$, we have $y' = -\tan y < 0$ and, therefore, $y \searrow$.

In fact, every solution y is decreasing (or increasing) throughout its domain. The conclusions are confirmed with Euler’s method:



We can match this ODE with a system:

$$\begin{cases} x' = 1, \\ y' = -\tan y. \end{cases}$$

Its solutions have these trajectories as solutions.

The directions of a parametric curves are its tangent vectors. Therefore, the directions of the solutions of the system of ODEs:

$$\begin{cases} x' = f(x, y), \\ y' = g(x, y), \end{cases}$$

are derived from the signs of the functions in the right-hand side:

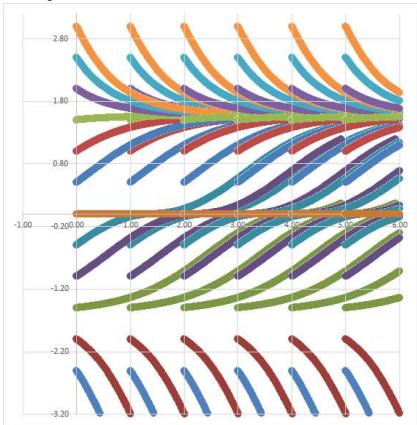
	$x' = f(x, y) < 0$	$x' = f(x, y) = 0$	$x' = f(x, y) > 0$
	←	•	→
$y' = g(x, y) < 0$	↓	↙	↘
$y' = g(x, y) = 0$	•	•	→
$y' = g(x, y) > 0$	↑	↖	↗

Example 4.7.2: more complex

Consider next:

$$y' = \cos y \cdot \sqrt{|y|}.$$

we demonstrated that the monotonicity of the solutions varies with the cosine:



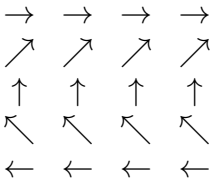
We can see that all solutions progress forward along the x -axis. What if we add variability of the direction of x ? Consider:

$$\begin{cases} x' = \sin y, \\ y' = \cos y. \end{cases}$$

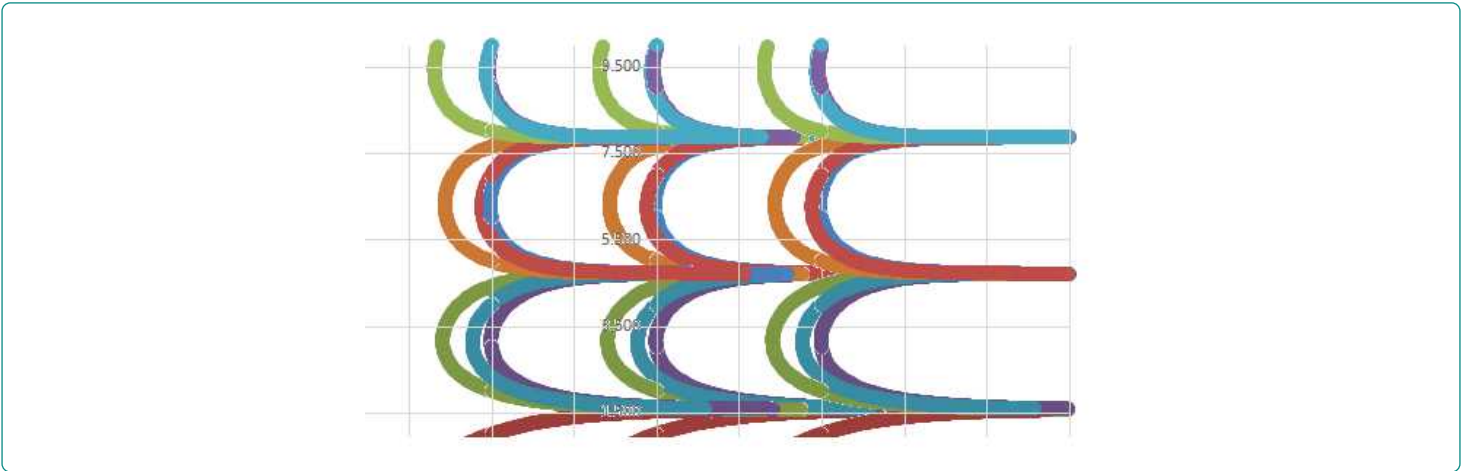
We conduct the “sign analysis” for both functions in the right-hand side:

y	x'	$x = x(t)$	path	y	y'	$y = y(t)$	path
2π	0			$3\pi/2$	0		
	−	$x \downarrow$	←←←←		−	$y \downarrow$	↓↓↓↓
π	0			$\pi/2$	0		
	+	$x \uparrow$	→→→→		+	$y \uparrow$	↑↑↑↑
0	0			$-\pi/2$	0		
	−	$x \downarrow$	←←←←				
$-\pi$	0						

Now put them together:



The results are confirmed with Euler’s method:



Example 4.7.3: trig

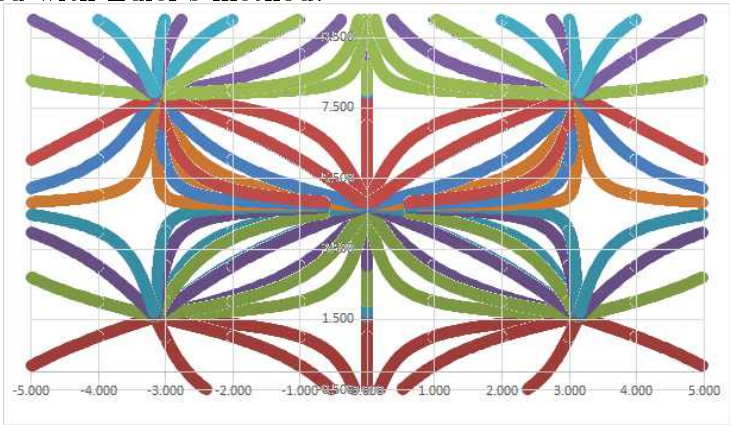
The next one:

$$\begin{cases} x' = \sin x, \\ y' = \cos y. \end{cases}$$

We conduct the “sign analysis” of the right-hand side:

	x	$-\pi$		π		2π		3π
y	$y' x'$	0	+	0	−	0	+	0
$3\pi/2$	0	●	→	●	←	●	→	●
	−	↓	↘	↓	↙	↓	↘	↓
$\pi/2$	0	●	→	●	←	●	→	●
	+	↑	↗	↑	↖	↑	↗	↑
$-\pi/2$	0	●	→	●	←	●	→	●
	−	↓	↘	↓	↙	↓	↘	↓
$-3\pi/2$	0	●	→	●	←	●	→	●

The results are confirmed with Euler’s method:

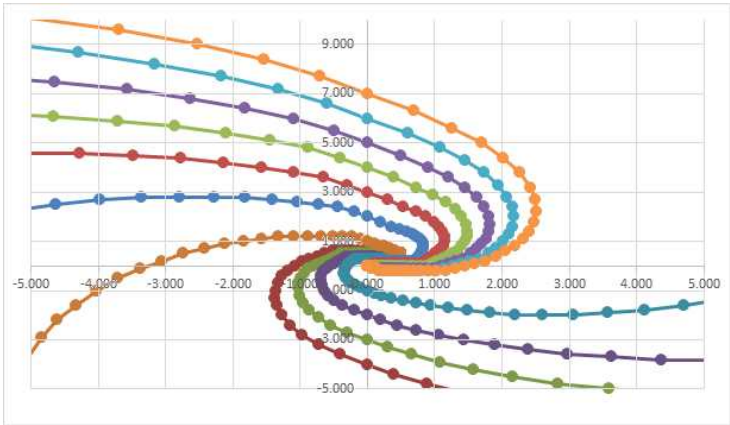


Example 4.7.4: discontinuous RHS

The next one is discontinuous:

$$x' = x - y, \quad y' = [x + y].$$

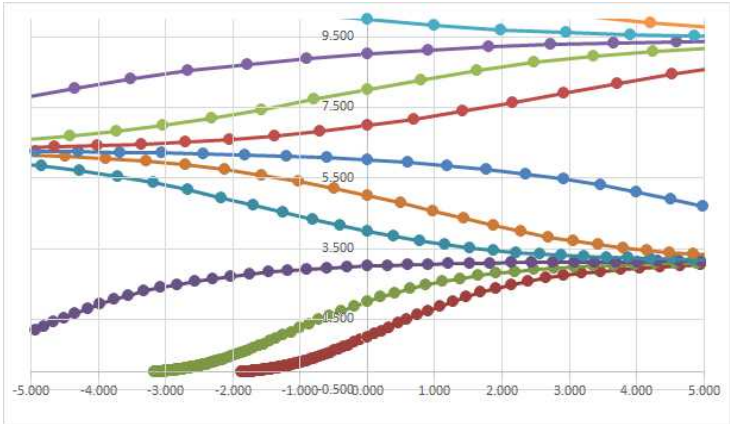
For Euler’s method we use the `FLOOR` function:



Exercise 4.7.5

Confirm the plot below by analyzing this system:

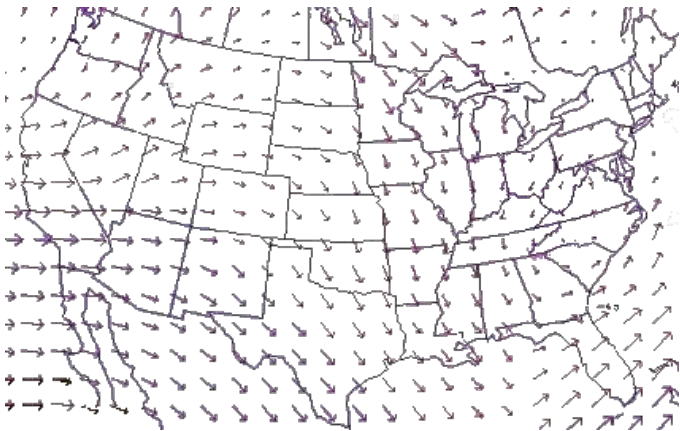
$$x' = y, \quad y' = \sin y \cdot y.$$



Suppose the system is *time-independent*,

$$x' = f(x,y), \quad y' = g(x,y).$$

Then it is thought of as a flow: liquid in a pond or the air over a surface of the Earth.



Then, the right-hand side is recognized as a *two-dimensional vector field*. In contrast to the 1-dimensional case with only three main possibilities, we will see a wide variety of behaviors around an equilibrium when the space of location is a plane.

With such a system, the qualitative analysis is much simpler. In fact, the ones above exhibit most of the possible patterns of *local* behavior.

We concentrate on what is going on in the vicinity of a given location $(x,y) = (a,b)$.

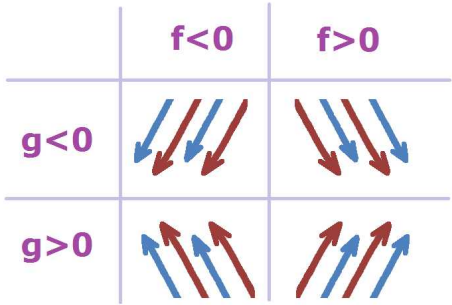
The first main possibility is

$$f(a,b) \neq 0 \text{ or } g(a,b) \neq 0,$$

which is equivalent to

$$F(a,b) = \langle f(a,b), g(a,b) \rangle \neq 0.$$

Then, from the continuity of f and g , we conclude that neither changes its sign in some disk D that contains (a,b) . Then, the solutions with paths located within D proceed in an about the same direction:



The behavior is “generic”.

More interesting behaviors are seen around a zero of F :

► $F(a,b) = 0 \implies (x,y) = (a,b)$ is a stationary solution (an equilibrium).

Then the pattern in the vicinity of the point, i.e., an open disk D , depends on whether this is a maximum of f or g , or a minimum, or neither. Some of the ideas come from dimension 1.

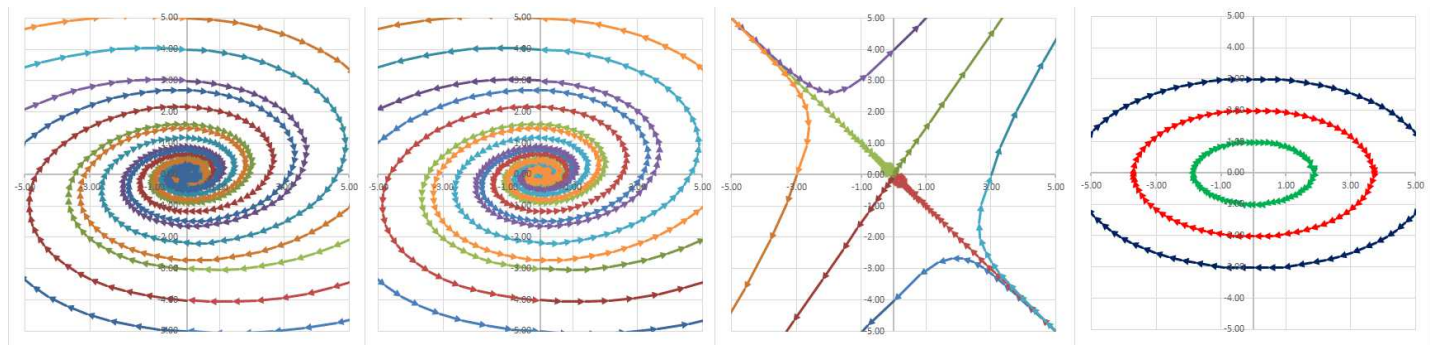
For example, a *stable equilibrium*,

$$y \rightarrow a \text{ as } x \rightarrow +\infty,$$

is a sink: flow in only. An *unstable equilibrium*,

$$y \rightarrow a \text{ as } x \rightarrow -\infty,$$

is a source: flow out only.



An *semi-stable equilibrium* could mean that some the solutions asymptotically approach the equilibrium and others do not. As you can see there are many more possibilities than in dimension 1.

4.8. The vector notation and linear systems

We can combine the two variables to form a point or a vector on the plane:

$$X = (x,y) \text{ or } X = \langle x,y \rangle .$$

We can also use the *column-vector* notation:

$$X = \begin{bmatrix} x \\ y \end{bmatrix} .$$

Next, the same happens to their derivatives as vectors:

$$\frac{\Delta X}{\Delta t} = \left\langle \frac{\Delta x}{\Delta t}, \frac{\Delta y}{\Delta t} \right\rangle = \begin{bmatrix} \frac{\Delta x}{\Delta t} \\ \frac{\Delta y}{\Delta t} \end{bmatrix}.$$

Then the setup we have been using for real-valued functions reappears:

$$\frac{\Delta X}{\Delta t} = F(X), \quad X(t_0) = X_0.$$

In this section, our main concern will be ODEs with respect to *derivatives*:

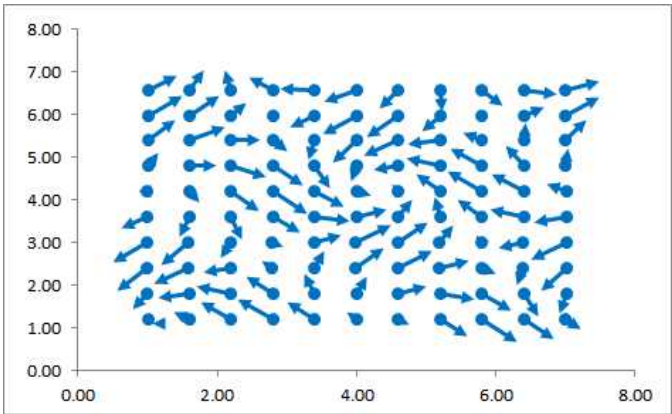
$$X' = \langle x', y' \rangle = \begin{bmatrix} x' \\ y' \end{bmatrix},$$

and

$$X' = F(X), \quad X(t_0) = X_0.$$

All the plotting, however, is done with the discrete ODEs.

The “phase space” \mathbf{R}^2 is the space of all possible locations. Then the position of a given particle is a function $X : \mathbf{R}^+ \rightarrow \mathbf{R}^2$ of time $t \geq 0$. Meanwhile, the dynamics of the particle is governed by the velocity of the flow, at each location, the same at every moment of time: the velocity of a particle if it happens to be at point X is $F(X)$.



Then either the next position is predicted to be $X + F(X)$ – that’s a discrete model – or $F(X)$ is just a tangent of the trajectory – that’s an ODE.

A vector field supplies a *direction to every location*, i.e., there is a vector attached to each point of the plane:

$$\text{point} \mapsto \text{vector}.$$

A *vector field* in dimension 2 is any function:

$$F : \mathbf{R}^2 \rightarrow \mathbf{R}^2.$$

Furthermore, one can think of a vector field as a *time-independent* ODE on the plane:

$$X' = F(X).$$

The corresponding IVP adds an initial condition:

$$X(t_0) = X_0.$$

Definition 4.8.1: solution of system of ODEs

A *solution* of a system of ODEs is a function u differentiable on an open interval I such that for every t in I we have:

$$X'(t) = F(X(t))$$

Definition 4.8.2: initial value problem

For a given system of ODEs and a given (t_0, X_0) , the *initial value problem*, or an IVP, is

$$X' = F(X), \quad X(t_0) = X_0$$

and its *solution* is a solution of the ODE that satisfies the *initial condition* above.

Definition 4.8.3: Euler solution

The *Euler solution* with increment $\Delta t > 0$ of the IVP:

$$X' = F(X), \quad X(t_0) = X_0;$$

is a sequence $\{X_n\}$ of points on the plane given by:

$$X_{n+1} = X_n + F(X_n) \cdot \Delta t$$

where $t_{n+1} = t_n + \Delta t$.

All the definitions above remain valid if we think of X as a location in an N -dimensional space \mathbf{R}^N . We now concentrate on *linear systems* (that may be acquired from non-linear ones via linearization). All the functions involved are linear and, therefore, differentiable; both existence and uniqueness are satisfied! This is the case of a linear function F . As such, it is given by a matrix and is evaluated via matrix multiplication:

$$F(X) = FX = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}.$$

It is written simply as

$$X' = FX.$$

The characteristics of this matrix – the determinant, the trace, and the eigenvalues – will help us classify such a system. However, the first observation is very simple: $X = 0$ is the equilibrium of the system. In fact what we’ve learned about systems of linear equations tells us the following.

Theorem 4.8.4: Equilibrium of Linear System

When $\det F \neq 0$, $X = 0$ is the only equilibrium of the system $X' = FX$; otherwise, there are infinitely many such points.

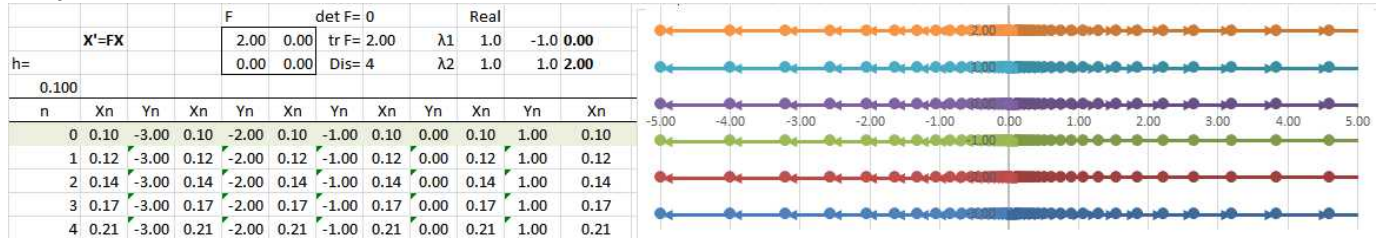
In other words, what matter is whether F , as a function, is or is not one-to-one.

Example 4.8.5: degenerate

The latter is the “degenerate” case such as the following. Let’s consider this very simple system of ODEs:

$$\begin{cases} x' = 2x \\ y' = 0 \end{cases} \implies \begin{cases} x = Ce^{2t} \\ y = K \end{cases}$$

It is easy to solve, one equation at a time. We have exponential growth on the x -axis and constant on the y -axis.

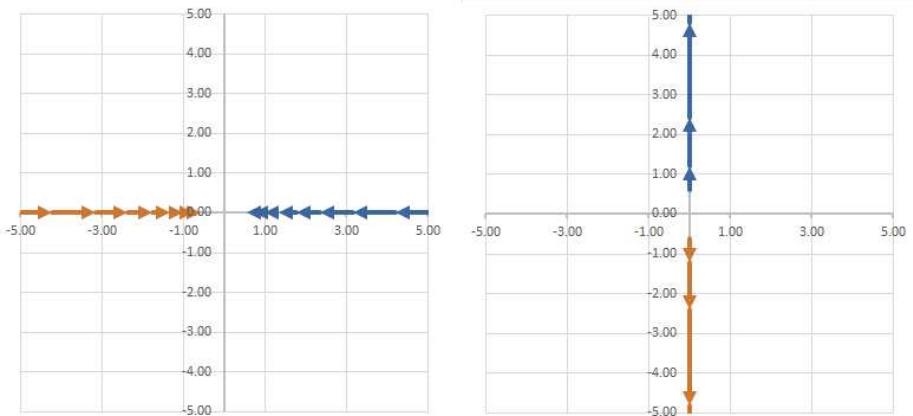


Example 4.8.6: saddle

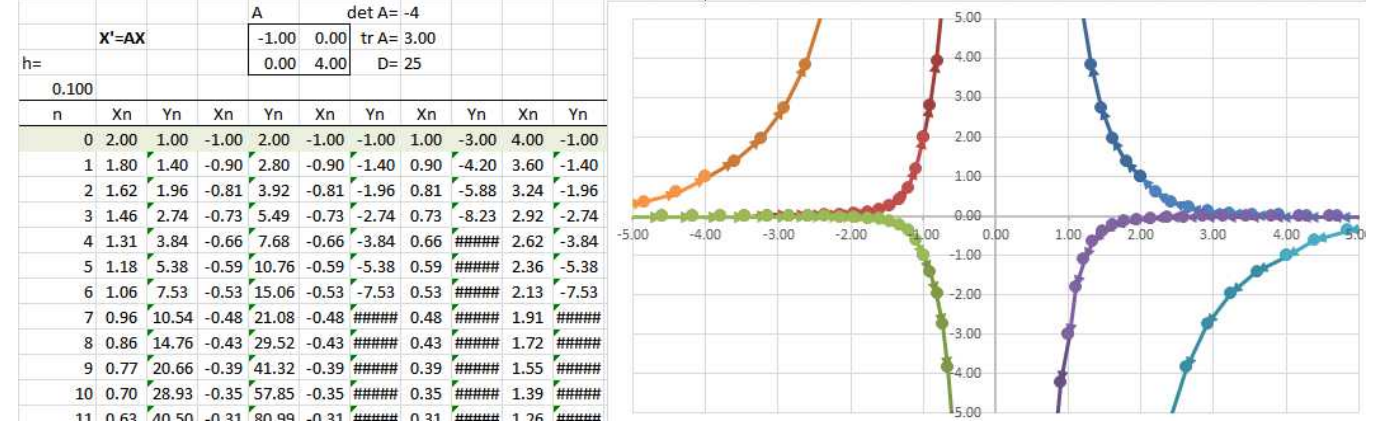
Let’s consider this system of ODEs:

$$\begin{cases} x' = -x \\ y' = 4y \end{cases} \implies \begin{cases} x = Ce^{-t} \\ y = Ke^{4t} \end{cases}$$

We solve it instantly because the two variables are fully separated. We can think of either of these two solutions of the two ODEs as a solution of the whole system that lives entirely within one of the two axes:



We have exponential growth on the x -axis and exponential decay on the y -axis. The rest of the solutions are seen to tend toward one of these. Since not all of the solutions go toward the origin, it is *unstable*.



This pattern is called a “saddle” because the curves look like the level curves of a function of two variables around a saddle point. Here, the matrix of F is diagonal:

$$F = \begin{bmatrix} -1 & 0 \\ 0 & 4 \end{bmatrix}.$$

Algebraically, we have:

$$X = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} Ce^{-t} \\ Ke^{4t} \end{bmatrix} = Ce^{-t} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + Ke^{4t} \begin{bmatrix} 0 \\ 1 \end{bmatrix} .$$

We have expressed the general solution as a linear combination of the two basis vectors!

Example 4.8.7: node

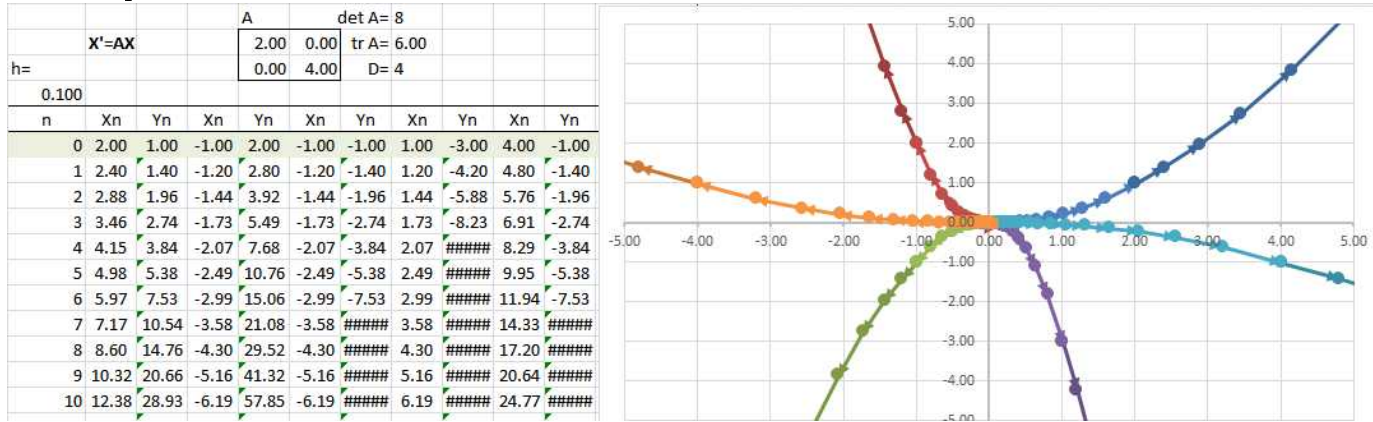
A slightly different system is:

$$\begin{cases} x' = 2x \\ y' = 4y \end{cases} \implies \begin{cases} x = Ce^{2t} \\ y = Ke^{4t} \end{cases}$$

Here,

$$F = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} .$$

Once again, either of these two solutions of the two ODEs is a solution of the whole system that lives entirely within one of the two axes (exponential growth on both of the axes) and the rest of the solutions are seen to tend toward one of these. The slight change to the system produces a very different pattern:

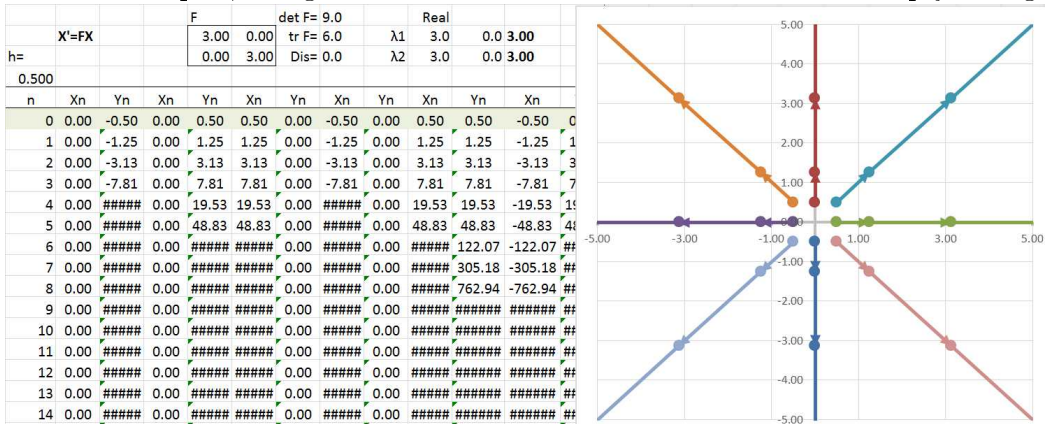


Algebraically, we have:

$$X = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} Ce^{2t} \\ Ke^{4t} \end{bmatrix} = Ce^{2t} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + Ke^{4t} \begin{bmatrix} 0 \\ 1 \end{bmatrix} .$$

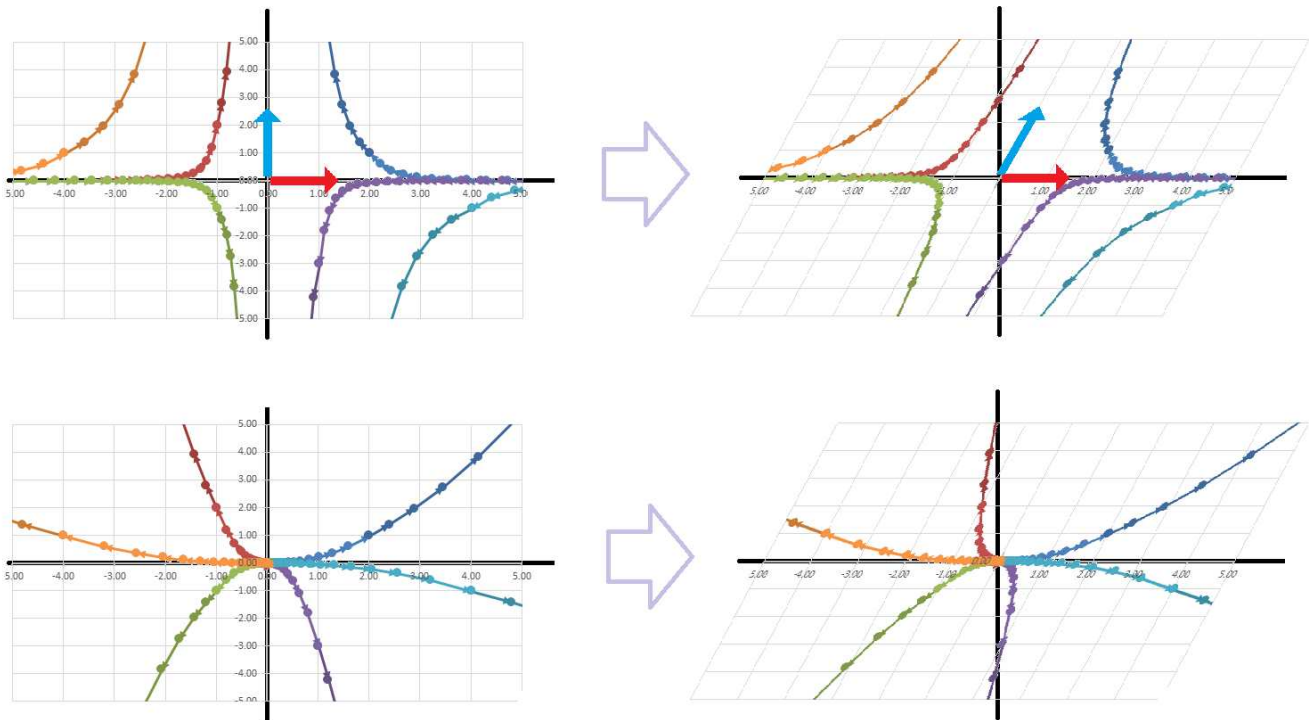
This is a linear combination of the two basis vectors with time-dependent weights.

The equilibrium is unstable but changing 2 and 4 to -2 and -4 will reverse the directions of the curves and make it stable. The exponential growth is faster along the y -axis; that is why the solutions appear to be tangent to the x -axis. In fact, eliminating t gives us $y = x^2$ and similar graphs. When the two coefficients are equal, the growth is identical and the solutions are simply straight lines:



What if the two variables aren't separated? The insight is that they can be – along the *eigenvectors* of the

matrix. Indeed, the basis vectors are the eigenvectors of these two diagonal matrices. Now just imagine that the two pictures are skewed:



Let’s look at them one at a time. The idea is uncomplicated: the system *within the eigenspace* is 1-dimensional. In other words, it is a single ODE and can be solved the usual way. This is how we find solutions. Every solution X that lies within the eigenspace, which is a line, is a t -dependent multiple of the eigenvector V :

$$\begin{aligned} X = rV &\implies X' = (rV)' = F(rV) = rFV = r\lambda V \\ &\implies r'V = \lambda rV \\ &\implies r' = \lambda r \\ &\implies r = e^{\lambda t} \end{aligned}$$

Theorem 4.8.8: Eigenspace Solutions

If λ is an eigenvalue and V a corresponding eigenvector of a matrix F , then

$$X = e^{\lambda t}V$$

is a solution of the linear system $X' = FX$.

Proof.

To verify, we substitute into the equation and use linearity (of both matrix multiplication and differentiation):

$$\begin{aligned} X = e^{\lambda t}V &\implies \\ \text{left-hand side: } X' &= (e^{\lambda t}V)' = (e^{\lambda t})'V = \lambda e^{\lambda t}V \\ \text{right-hand side: } FX &= F(e^{\lambda t}V) = e^{\lambda t}FV = e^{\lambda t}\lambda V \end{aligned}$$

The eigenvalue can be *complex*!

The second idea is to try to express the solution of the general linear system as a linear combination of two solutions found this way.

Theorem 4.8.9: Representation In Terms of Eigensolutions

Suppose V_1 and V_2 are two eigenvectors of a matrix F that correspond to two eigenvalues λ_1 and λ_2 . Suppose also that the eigenvectors aren't multiples of each other. Then all solutions of the linear system $X' = FX$ are given as linear combinations of non-trivial solutions within the eigenspaces:

$$X = Ce^{\lambda_1 t}V_1 + Ke^{\lambda_2 t}V_2,$$

with real coefficients C and K .

Proof.

Since these solutions cover the whole plane, the conclusion follows from the uniqueness property.

Definition 4.8.10: characteristic solution

For a given eigenvector V with eigenvalue λ , we will call $e^{\lambda t}V$ a *characteristic solution*.

Exercise 4.8.11

Show that when all eigenvectors are multiples of each other, the formula won't give us *all* the solutions.

4.9. Classification of linear systems

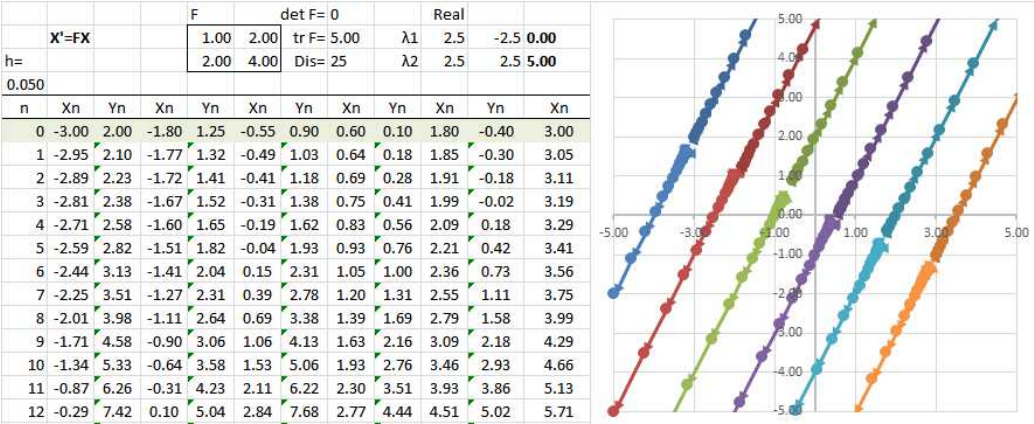
We consider a few systems with *non-diagonal* matrices. The computations of the eigenvalues and eigenvectors come from Chapter 3.

Example 4.9.1: degenerate

Let's consider a more general system of ODEs:

$$\begin{cases} x' = x + 2y, \\ y' = 2x + 4y, \end{cases} \implies F = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}.$$

Euler's method shows the following solutions:



It appears that the system has (exponential) growth in one direction and constant in another. What are those directions? Linear algebra helps.

First, the determinant is zero:

$$\det F = \det \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} = 1 \cdot 4 - 2 \cdot 2 = 0.$$

That’s why there is a whole line of points X with $FX = 0$. These are stationary points. To find them, we solve this equation:

$$\begin{cases} x & +2y & = 0, \\ 2x & +4y & = 0, \end{cases} \implies x = -2y.$$

We have, then, eigenvectors corresponding to the zero eigenvalue $\lambda_1 = 0$:

$$V_1 = \begin{bmatrix} 2 \\ -1 \end{bmatrix} \implies FV_1 = 0.$$

So, second, there is only one non-zero eigenvalue:

$$\det(F - \lambda I) = \det \begin{bmatrix} 1 - \lambda & 2 \\ 2 & 4 - \lambda \end{bmatrix} = \lambda^2 - 5\lambda = \lambda(\lambda - 5).$$

Let’s find the eigenvectors for $\lambda_2 = 5$. We solve the equation:

$$FV = \lambda_2 V,$$

as follows:

$$FV = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 5 \begin{bmatrix} x \\ y \end{bmatrix}.$$

This gives as the following system of linear equations:

$$\begin{cases} x & +2y & = 5x, \\ 2x & +4y & = 5y, \end{cases} \implies \begin{cases} -4x & +2y & = 0, \\ 2x & -y & = 0, \end{cases} \implies y = 2x.$$

This line is the eigenspace. We choose the eigenvector to be:

$$V_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Every solution starts off the line $y = -x/2$ and continues along this vector. It is a linear combination of the two eigenvectors:

$$X = CV_1 + KV_2 = C \begin{bmatrix} 2 \\ -1 \end{bmatrix} + Ke^{5t} \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Exercise 4.9.2

Find the line of stationary solutions.

Example 4.9.3: saddle

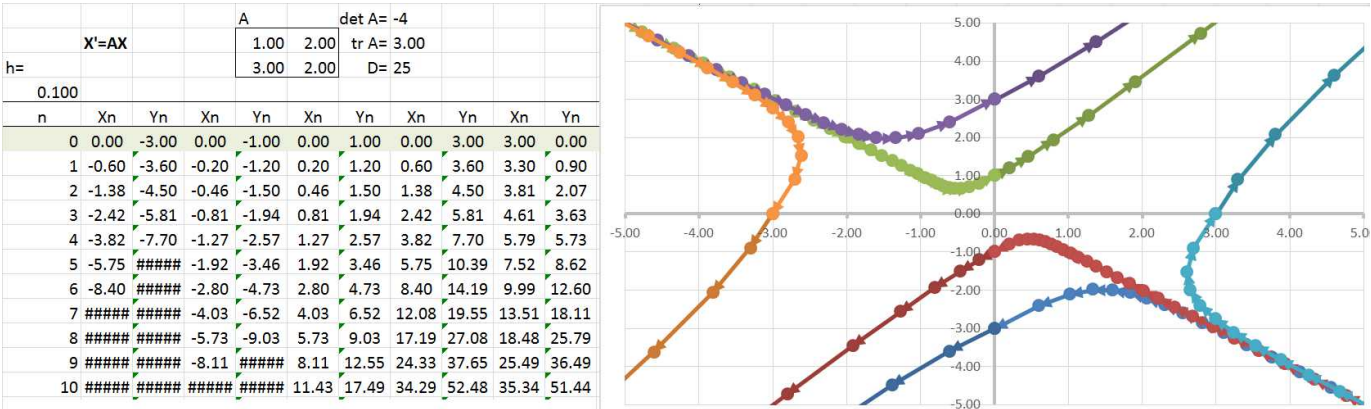
Let’s consider this system of ODEs:

$$\begin{cases} x' & = x & +2y, \\ y' & = 3x & +2y. \end{cases}$$

Here, the matrix of F is not diagonal:

$$F = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}.$$

Euler’s method shows the following:



The two lines the solutions appear to converge to are the eigenspaces. Let’s find them:

$$\det(F - \lambda I) = \det \begin{bmatrix} 1 - \lambda & 2 \\ 3 & 2 - \lambda \end{bmatrix} = \lambda^2 - 3\lambda - 4.$$

Therefore, the eigenvalues are

$$\lambda_1 = -1, \lambda_2 = 4.$$

Now we find the eigenvectors. We solve the two equations:

$$FV_k = \lambda_k V_k, \quad k = 1, 2.$$

The first:

$$FV_1 = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = -1 \begin{bmatrix} x \\ y \end{bmatrix}.$$

This gives as the following system of linear equations:

$$\begin{cases} x + 2y = -x, \\ 3x + 2y = -y, \end{cases} \implies \begin{cases} 2x + 2y = 0, \\ 3x + 3y = 0, \end{cases} \implies x = -y.$$

We choose

$$V_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Every solution within this eigenspace (the line $y = -x$) is a multiple of this characteristic solution:

$$X_1 = e^{\lambda_1 t} V_1 = e^{-t} \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

The second eigenvalue:

$$FV_2 = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 4 \begin{bmatrix} x \\ y \end{bmatrix}.$$

We have the following system:

$$\begin{cases} x + 2y = 4x, \\ 3x + 2y = 4y, \end{cases} \implies \begin{cases} -3x + 2y = 0, \\ 3x - 2y = 0, \end{cases} \implies x = 2y/3.$$

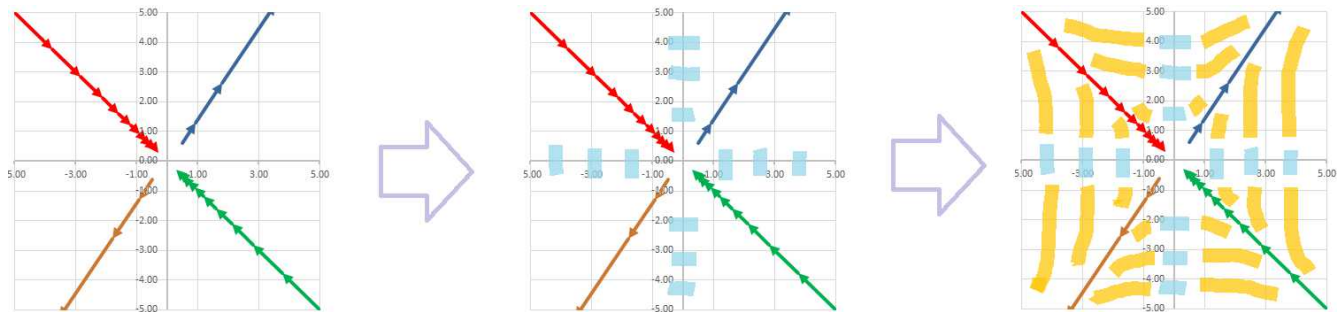
We choose

$$V_2 = \begin{bmatrix} 1 \\ 3/2 \end{bmatrix}.$$

Every solution within this eigenspace (the line $y = 3x/2$) is a multiple of this characteristic solution:

$$X_2 = e^{\lambda_2 t} V_2 = e^{4t} \begin{bmatrix} 1 \\ 3/2 \end{bmatrix}.$$

The two solutions X_1 and X_2 , as well as $-X_1$ and $-X_2$, are shown below:



The general solution is a linear combination of these two basic solutions:

$$X = Ce^{\lambda_1 t}V_1 + Ke^{\lambda_2 t}V_2 = Ce^{-t} \begin{bmatrix} 1 \\ -1 \end{bmatrix} + Ke^{4t} \begin{bmatrix} 1 \\ 3/2 \end{bmatrix} = \begin{bmatrix} Ce^{-t} + Ke^{4t} \\ -Ce^{-t} + 3/2Ke^{4t} \end{bmatrix},$$

i.e.,

$$\begin{cases} x &= Ce^{-t} + Ke^{4t}, \\ y &= -Ce^{-t} + 3/2Ke^{4t}. \end{cases}$$

The equilibrium is unstable.

Example 4.9.4: node

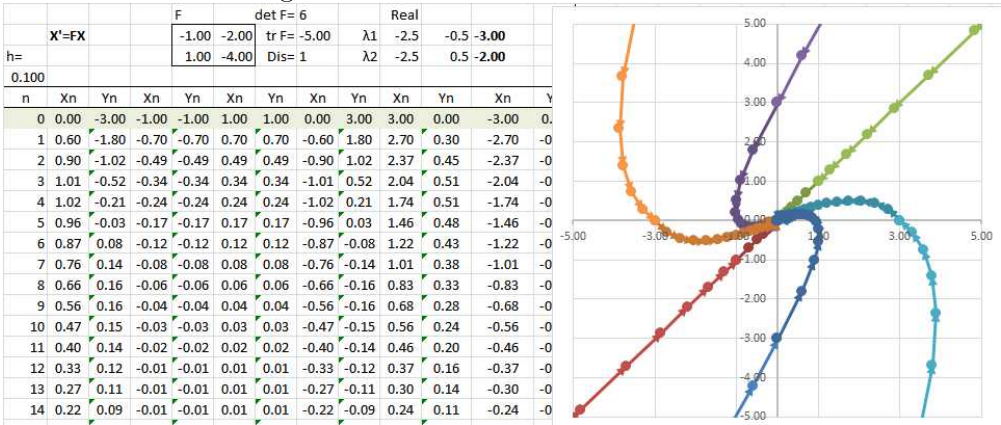
Let’s consider this system of ODEs:

$$\begin{cases} x' &= -x - 2y, \\ y' &= x - 4y. \end{cases}$$

Here, the matrix of F is not diagonal:

$$F = \begin{bmatrix} -1 & -2 \\ 1 & -4 \end{bmatrix}.$$

Euler’s method shows the following:



The analysis starts with the characteristic polynomial:

$$\det(F - \lambda I) = \det \begin{bmatrix} -1 - \lambda & -2 \\ 1 & -4 - \lambda \end{bmatrix} = \lambda^2 - 5\lambda + 6.$$

Therefore, the eigenvalues are

$$\lambda_1 = -3, \lambda_2 = -2.$$

To find the eigenvectors, we solve the two equations:

$$FV_k = \lambda_k V_k, \quad k = 1, 2.$$

The first:

$$FV_1 = \begin{bmatrix} -1 & -2 \\ 1 & -4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = -1 \begin{bmatrix} x \\ y \end{bmatrix}.$$

This gives as the following system of linear equations:

$$\begin{cases} -x & -2y & = -3x, \\ x & -4y & = -3y, \end{cases} \implies \begin{cases} 2x & -2y & = 0, \\ x & -y & = 0, \end{cases} \implies x = y.$$

We choose

$$V_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Every solution within this eigenspace (the line $y = x$) is a multiple of this characteristic solution:

$$X_1 = e^{\lambda_1 t} V_1 = e^{-3t} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

The second eigenvalue:

$$FV_2 = \begin{bmatrix} -1 & -2 \\ 1 & -4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = -2 \begin{bmatrix} x \\ y \end{bmatrix}.$$

We have the following system:

$$\begin{cases} -x & -2y & = -2x, \\ x & -4y & = -2y, \end{cases} \implies \begin{cases} x & -2y & = 0, \\ x & -2y & = 0, \end{cases} \implies x = 2y.$$

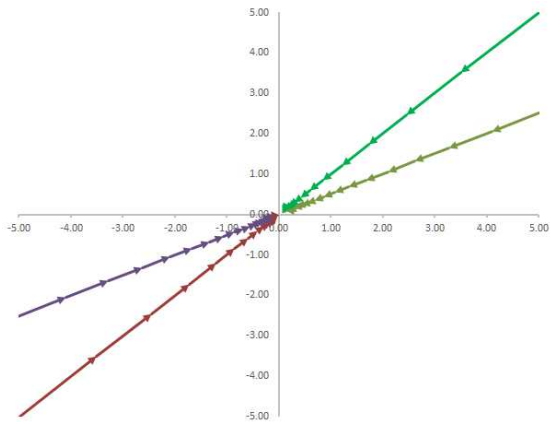
We choose

$$V_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

Every solution within this eigenspace (the line $y = x/2$) is a multiple of this this characteristic solution:

$$X_2 = e^{\lambda_2 t} V_2 = e^{-2t} \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

The two solutions X_1 and X_2 , as well as $-X_1$ and $-X_2$, are shown below:



The general solution is a linear combination of these two basic solutions:

$$X = Ce^{\lambda_1 t} V_1 + Ke^{\lambda_2 t} V_2 = Ce^{-3t} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + Ke^{-2t} \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

The equilibrium is stable.

Definition 4.9.5: stable node

For a linear system $X' = FX$, the equilibrium solution $X_0 = 0$ is called a *stable node* if every other solution X satisfies:

$$X(t) \rightarrow 0 \text{ as } t \rightarrow +\infty \text{ and } ||X(t)|| \rightarrow \infty \text{ as } t \rightarrow -\infty;$$

and an *unstable node* if

$$X(t) \rightarrow 0 \text{ as } t \rightarrow -\infty \text{ and } ||X(t)|| \rightarrow \infty \text{ as } t \rightarrow +\infty;$$

provided no X makes a full rotation around 0.

Definition 4.9.6: saddle

For a linear system $X' = FX$, the equilibrium solution $X_0 = 0$ is called a *saddle* if it has solutions X that satisfy:

$$||X(t)|| \rightarrow \infty \text{ as } t \rightarrow \pm\infty .$$

Theorem 4.9.7: Classification of Linear Systems I

Suppose matrix F has two real eigenvalues λ_1 and λ_2 . Then, we have:

- If λ_1 and λ_2 have the same sign, the system $X' = FX$ has a *node*, stable when this sign is negative and *unstable* when this sign is positive.
- If λ_1 and λ_2 have the opposite signs, the system $X' = FX$ has a *saddle*.

Proof.

The stability is seen in either of the two characteristic solutions, as $t \rightarrow +\infty$:

$$||X|| = ||e^{\lambda t}V|| = e^{\lambda t} \cdot ||V|| \rightarrow \begin{cases} \infty & \text{if } \lambda > 0, \\ 0 & \text{if } \lambda < 0. \end{cases}$$

According to the last theorem, we have a linear combination of the two characteristic solutions. Then, in the former case, we have one or the other pattern, and in the latter, both. There can be no rotation because no solution can intersect an eigenspace, according to the uniqueness property.

4.10. Classification of linear systems, continued

What if the eigenvalues are *complex*?

Recall that the characteristic polynomial of matrix F is

$$\chi(\lambda) = \det(F - \lambda I) = \lambda^2 - \operatorname{tr} F \cdot \lambda + \det F .$$

The discriminant of this quadratic polynomial is

$$D = (\operatorname{tr} F)^2 - 4 \det F .$$

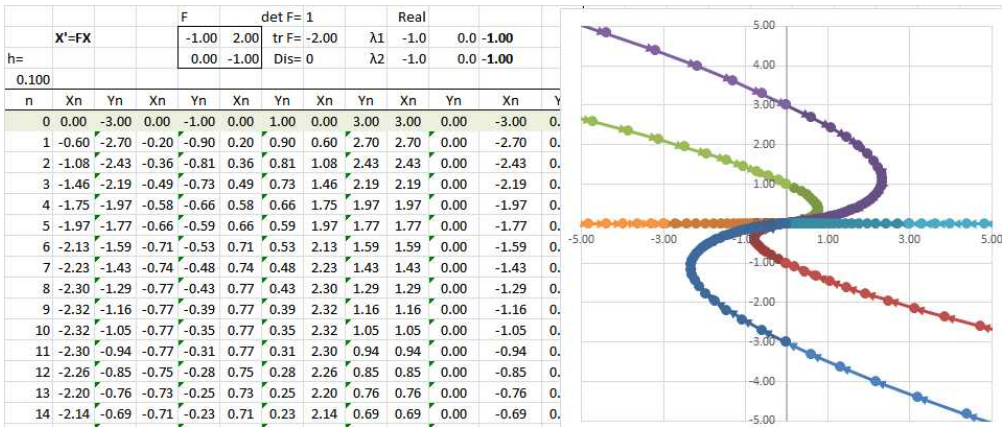
When $D > 0$, we have two distinct real eigenvalues, the case addressed in the last section. We are faced with complex eigenvalues whenever $D < 0$. The transitional case is $D = 0$.

Example 4.10.1: improper node

In contrast to the last example, a node may be produced by a matrix with repeated (and, therefore, real) eigenvalues:

$$F = \begin{bmatrix} -1 & 2 \\ 0 & -1 \end{bmatrix} .$$

Euler’s method shows the following:



The analysis starts with the characteristic polynomial:

$$\det(F - \lambda I) = \det \begin{bmatrix} -1 - \lambda & 2 \\ 0 & -1 - \lambda \end{bmatrix} = (-1 - \lambda)^2.$$

Therefore, the eigenvalues are

$$\lambda_1 = \lambda_2 = -1.$$

The only eigenvectors are horizontal. The solution is given by

$$X = Ce^{-t} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + K \left(te^{-t} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + e^{-t} \begin{bmatrix} ? \\ 1 \end{bmatrix} \right).$$

Exercise 4.10.2

Finish the computation in the example.

When $D < 0$, the eigenvalues are complex! Therefore, there are no eigenvectors (not real ones anyway). Does the system $X' = FX$ even have solutions? The theorem about characteristic solutions says yes, they are certain exponential functions...

Example 4.10.3: center

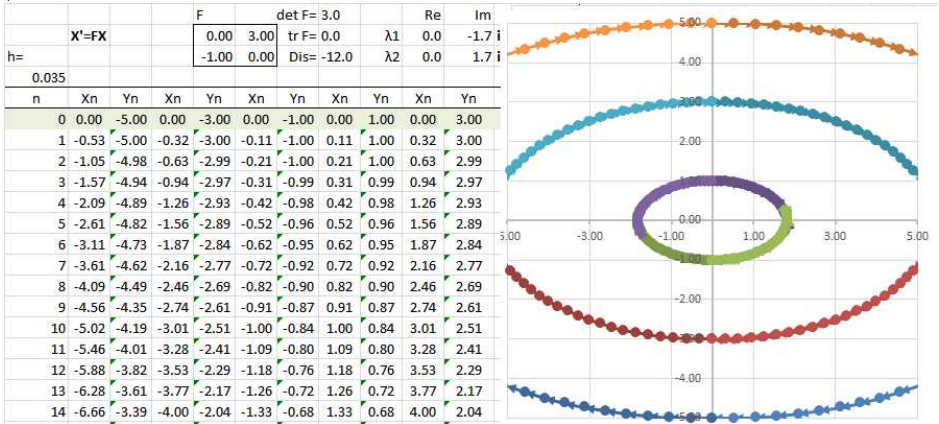
Consider

$$\begin{cases} x' &= y, \\ y' &= -x. \end{cases}$$

We already know that the solution is found by substitution:

$$x'' = (x')' = y' = -x.$$

Therefore the solutions are the linear combinations of $\sin t$ and $\cos t$. The result is confirmed with Euler's method (with a limited number of step to prevent the approximations to spiral out):



According to the theory above, the solutions are supposed to be exponential rather than trigonometric. But the latter are just exponential functions with imaginary exponents.

Let's make this specific; we have

$$F = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},$$

and the characteristic polynomial,

$$\chi(\lambda) = \lambda^2 + 1,$$

has these complex roots: $\lambda_{1,2} = \pm i$.

To find the first eigenvector, we solve:

$$FV_1 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = i \begin{bmatrix} x \\ y \end{bmatrix}.$$

This gives as the following system of linear equations:

$$\begin{cases} y = ix \\ -x = iy \end{cases} \implies y = ix.$$

We choose a *complex* eigenvector:

$$V_1 = \begin{bmatrix} 1 \\ i \end{bmatrix},$$

and similarly:

$$V_2 = \begin{bmatrix} 1 \\ -i \end{bmatrix},$$

The general solution is a linear combination – over the complex numbers – of these two characteristic solutions:

$$X = Ce^{\lambda_1 t}V_1 + Ke^{\lambda_2 t}V_2 = Ce^{it} \begin{bmatrix} 1 \\ i \end{bmatrix} + Ke^{-it} \begin{bmatrix} 1 \\ -i \end{bmatrix}.$$

The problem is solved! ...in the complex domain. What is the *real part*?

Let $K = 0$. Then the solution is:

$$X = Ce^{it} \begin{bmatrix} 1 \\ i \end{bmatrix} = C \begin{bmatrix} e^{it} \\ ie^{it} \end{bmatrix} = C \begin{bmatrix} \cos t + i \sin t \\ i(\cos t + i \sin t) \end{bmatrix} = C \begin{bmatrix} \cos t + i \sin t \\ -\sin t + i \cos t \end{bmatrix}.$$

Its real part is:

$$\text{Re } X = C \begin{bmatrix} \cos t \\ -\sin t \end{bmatrix}.$$

These are all the circles.

Example 4.10.4: focus

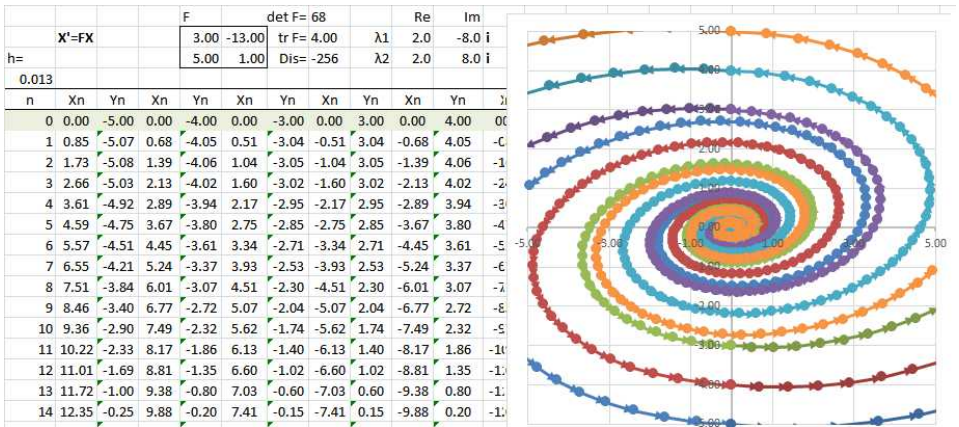
Let's consider a more complex system ODEs:

$$\begin{cases} x' = 3x - 13y, \\ y' = 5x + y. \end{cases}$$

Here, the matrix of F is not diagonal:

$$F = \begin{bmatrix} 3 & -13 \\ 5 & 1 \end{bmatrix}.$$

Euler's method shows the following:



The analysis starts with the characteristic polynomial:

$$\chi(\lambda) = \det(F - \lambda I) = \det \begin{bmatrix} 3 - \lambda & -13 \\ 5 & 1 - \lambda \end{bmatrix} = \lambda^2 - 4\lambda + 68.$$

Therefore, the eigenvalues are

$$\lambda_{1,2} = 2 \pm 8i.$$

Now we find the eigenvectors. We solve the two equations:

$$FV_k = \lambda_k V_k, \quad k = 1, 2.$$

The first:

$$FV_1 = \begin{bmatrix} 3 & -13 \\ 5 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = (2 + 8i) \begin{bmatrix} x \\ y \end{bmatrix}.$$

This gives as the following system of linear equations:

$$\begin{cases} 3x - 13y = (2 + 8i)x \\ 5x + y = (2 + 8i)y \end{cases} \implies \begin{cases} (1 - 8i)x - 13y = 0 \\ 5x + (-1 - 8i)y = 0 \end{cases} \implies x = \frac{1 + 8i}{5}y.$$

We choose

$$V_1 = \begin{bmatrix} 1 + 8i \\ 5 \end{bmatrix}.$$

The second eigenvalue:

$$FV_2 = \begin{bmatrix} 3 & -13 \\ 5 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = (2 - 8i) \begin{bmatrix} x \\ y \end{bmatrix}.$$

We have the following system:

$$\begin{cases} 3x - 13y = (2 - 8i)x \\ 5x + y = (2 - 8i)y \end{cases} \implies \begin{cases} (1 + 8i)x - 13y = 0 \\ 5x + (-1 + 8i)y = 0 \end{cases} \implies x = \frac{1 - 8i}{5}y.$$

We choose

$$V_2 = \begin{bmatrix} 1 - 8i \\ 5 \end{bmatrix}.$$

The general complex solution is a linear combination of the two characteristic solutions:

$$Z = Ce^{\lambda_1 t} V_1 + Ke^{\lambda_2 t} V_2 = Ce^{(2+8i)t} \begin{bmatrix} 1 + 8i \\ 5 \end{bmatrix} + Ke^{(2-8i)t} \begin{bmatrix} 1 - 8i \\ 5 \end{bmatrix}.$$

Let's now examine a simple *real* solution. We let $C = 1$ and $K = 0$:

$$\begin{aligned} X = \operatorname{Re} Z &= \operatorname{Re} e^{(2+8i)t} \begin{bmatrix} 1+8i \\ 5 \end{bmatrix} \\ &= e^{2t} \operatorname{Re} e^{8it} \begin{bmatrix} 1+8i \\ 5 \end{bmatrix} \\ &= e^{2t} \operatorname{Re}(\cos 8t + i \sin 8t) \begin{bmatrix} 1+8i \\ 5 \end{bmatrix} \\ &= e^{2t} \operatorname{Re} \begin{bmatrix} (\cos 8t + i \sin 8t)(1+8i) \\ (\cos 8t + i \sin 8t)5 \end{bmatrix} \\ &= e^{2t} \operatorname{Re} \begin{bmatrix} \cos 8t + i \sin 8t + 8i \cos 8t - 8 \sin 8t \\ 5 \cos 8t + i 5 \sin 8t \end{bmatrix} \\ &= e^{2t} \begin{bmatrix} \cos 8t - 8 \sin 8t \\ 5 \cos 8t \end{bmatrix}. \end{aligned}$$

Plotting this parametric curve confirms Euler's method result:

Definition 4.10.5: stable and unstable focus

For a linear system $X' = FX$, the equilibrium solution $X_0 = 0$ is called a *stable focus* if every other solution X satisfies:

$$X(t) \rightarrow 0 \text{ as } t \rightarrow +\infty \text{ and } \|X(t)\| \rightarrow \infty \text{ as } t \rightarrow -\infty;$$

and an *unstable focus* if

$$X(t) \rightarrow 0 \text{ as } t \rightarrow -\infty \text{ and } \|X(t)\| \rightarrow \infty \text{ as } t \rightarrow +\infty;$$

provided every such X makes a full rotation around 0.

Definition 4.10.6: center

For a linear system $X' = FX$, the equilibrium solution $X_0 = 0$ is called a *center* if all solutions are cycles.

Theorem 4.10.7: Classification of Linear Systems II

Suppose matrix F has two complex conjugate eigenvalues λ_1 and λ_2 . Then we have:

- If the real part of λ_1 and λ_2 is non-zero, the system $X' = FX$ has a focus, stable when this sign of this number is negative and unstable when this sign is positive.
- If the real part of λ_1 and λ_2 is zero, the system $X' = FX$ has a center.

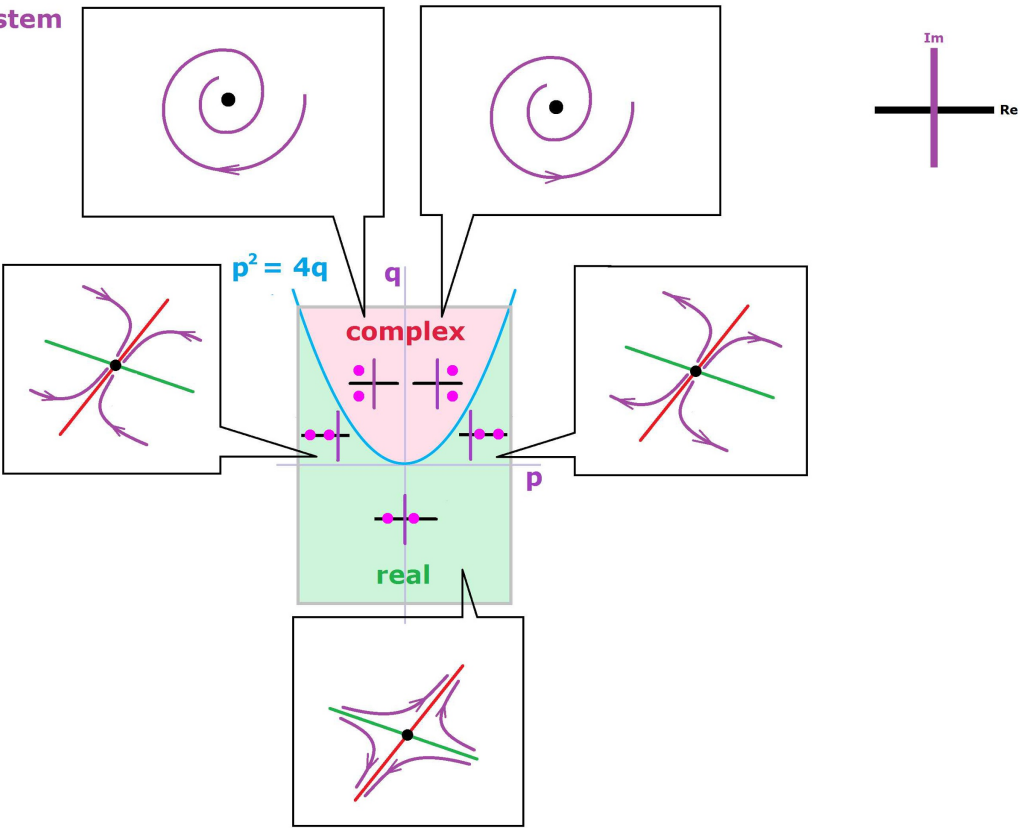
Proof.

The stability is seen in either of the two characteristic solutions, as $t \rightarrow +\infty$:

$$\|X\| = \|e^{\lambda t}V\| = \|e^{(a+bi)t}V\| = e^{at}|\cos bt + i \sin bt| \cdot \|V\| = e^{at}\|V\| \rightarrow \begin{cases} \infty & \text{if } a > 0, \\ 0 & \text{if } a < 0. \end{cases}$$

The combination of the two classification theorems is illustrated below:

The eigenvalues of the system $X'=AX$ are the roots of $x^2 + px + q = 0$, where $p = -\text{tr } A$, $q = \det A$.



Thereby we complete the sequence: elementary algebra \longrightarrow matrix algebra \longrightarrow linear differential equations. To summarize, in order to classify a system of linear ODEs $X' = FX$, where F is a 2×2 matrix and X is a vector on the plane, we classify F according to its eigenvalues and visualize how the locations of these two numbers in the complex plane indicate very different behaviors of the trajectories. (The missing patterns are better illustrated dynamically, as the exact “moments” when one pattern transitions into another.)

Exercise 4.10.8

Point out on the complex plane the locations of the center and the improper node.

Exercise 4.10.9

How likely would a given system fall into each of these five categories? What about the center and the improper node?

Exercise 4.10.10

What parameters determine the clockwise vs. counter-clockwise behavior?

Example 4.10.11: predator-prey

Let’s classify the equilibria of the predator-prey model – via linearization. Our non-linear system is given by the following:

$$\begin{cases} x' = \alpha x - \beta xy, \\ y' = \delta xy - \gamma y, \end{cases}$$

with non-negative coefficients $\alpha, \beta, \delta, \gamma$. In other words, we have

$$X' = G(X), \text{ with } G(x, y) = (\alpha x - \beta xy, \delta xy - \gamma y).$$

The *Jacobian* of G is

$$G'(x, y) = \begin{bmatrix} \frac{\partial}{\partial x}(\alpha x - \beta xy) & \frac{\partial}{\partial y}(\alpha x - \beta xy) \\ \frac{\partial}{\partial x}(\delta xy - \gamma y) & \frac{\partial}{\partial y}(\delta xy - \gamma y) \end{bmatrix} = \begin{bmatrix} \alpha - \beta y & -\beta x \\ \delta y & \delta x - \gamma \end{bmatrix}.$$

The matrix depends on (x, y) because the system is non-linear. By fixing locations $X = A$, we create linear vector ODEs:

$$X' = G'(A)X.$$

First, we consider the zero equilibrium,

$$x = 0, \ y = 0.$$

Here,

$$F = G'(0, 0) = \begin{bmatrix} \alpha - \beta \cdot 0 & -\beta \cdot 0 \\ \delta \cdot 0 & \delta \cdot 0 - \gamma \end{bmatrix} = \begin{bmatrix} \alpha & 0 \\ 0 & -\gamma \end{bmatrix}.$$

The eigenvalues are found by solving the following equation:

$$\det \begin{bmatrix} \alpha - \lambda & 0 \\ 0 & -\gamma - \lambda \end{bmatrix} = (\alpha - \lambda)(-\gamma - \lambda) = 0.$$

Therefore,

$$\lambda_1 = \alpha, \ \lambda_2 = -\gamma.$$

We have two real eigenvalues of opposite signs. This is a *saddle*! Indeed, around this point the foxes decline while the rabbits increase in numbers.

The main equilibrium is

$$x = \frac{\gamma}{\delta}, \ y = \frac{\alpha}{\beta}.$$

Here,

$$F = G' \left(\frac{\gamma}{\delta}, \frac{\alpha}{\beta} \right) = \begin{bmatrix} \alpha - \beta \cdot \frac{\alpha}{\beta} & -\beta \cdot \frac{\gamma}{\delta} \\ \delta \cdot \frac{\alpha}{\beta} & \delta \cdot \frac{\gamma}{\delta} - \gamma \end{bmatrix} = \begin{bmatrix} 0 & -\frac{\beta\gamma}{\delta} \\ \frac{\delta\alpha}{\beta} & 0 \end{bmatrix}.$$

The eigenvalues are found by solving the following equation:

$$\det \begin{bmatrix} -\lambda & -\frac{\beta\gamma}{\delta} \\ \frac{\delta\alpha}{\beta} & -\lambda \end{bmatrix} = \lambda^2 + \alpha\gamma = 0.$$

Therefore,

$$\lambda_1 = \sqrt{\alpha\gamma}i, \lambda_2 = -\sqrt{\alpha\gamma}i.$$

We have two purely imaginary eigenvalues. This is a *center*! Indeed, around this point we have a cyclic behavior.

The results match our previous analysis.

Chapter 5: Applications of ODEs

Contents

5.1 Vector-valued forms	306
5.2 The pursuit curves	307
5.3 ODEs of second order as systems	312
5.4 Vector ODEs of second order: a double spring	315
5.5 A pendulum	321
5.6 Planetary motion	324
5.7 The two- and three-body problems	330
5.8 A cannon is fired...	336
5.9 Boundary value problems	339

5.1. Vector-valued forms

the example of parametric curves – and especially motion in space – suggests that we may need the domain of these functions to be multi-dimensional. We saw a function defined, just like a 0-form, at the nodes of the cell decomposition of the line, but with values in \mathbf{R}^2 , unlike a 0-form.

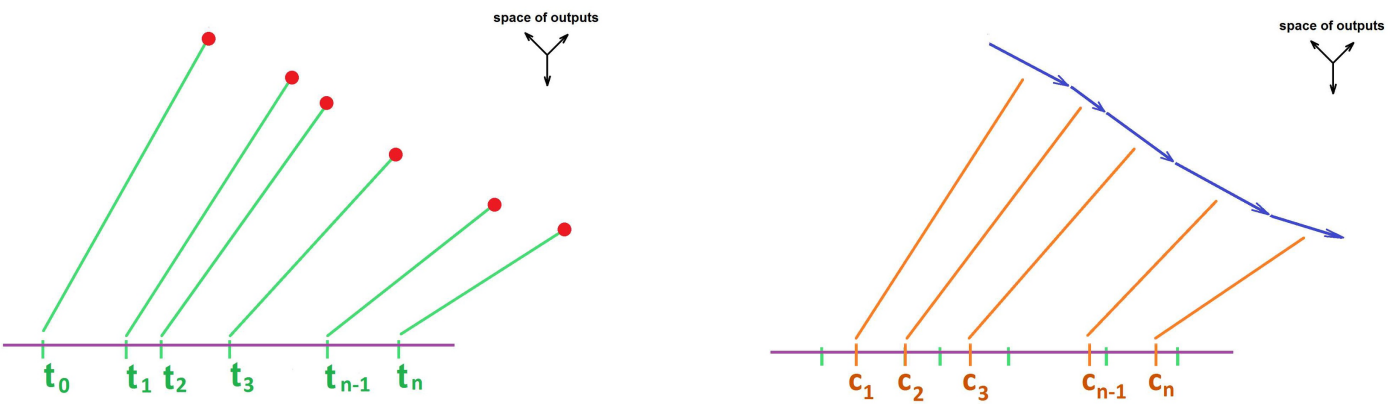
This is a generalization of the last definition.

Definition 5.1.1: vector-valued discrete form

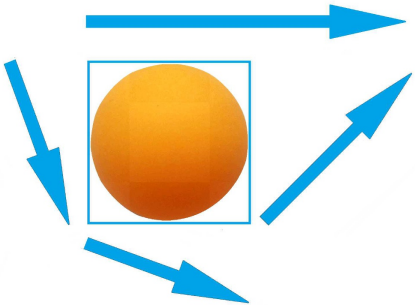
Suppose n and m are given. Then a *vector-valued discrete form F of degree k* , or simply a k -form, is a function defined on k -cells of \mathbf{R}^n with values in \mathbf{R}^m .

As you can see, we will be using capital letters for vector-valued forms in accordance with our convention. Note that discrete forms do *not* exactly match our list of functions: numerical functions, parametric curves, vector fields, and functions of several variables. From the same domain, we pick cells of different dimensions producing forms of different degrees.

This is an illustration of two vector-valued forms: a 0-form and a 1-form (for the latter, the vectors have to be moved to put the starting points at the origin); i.e., $n = 1$ and $m = 2$:



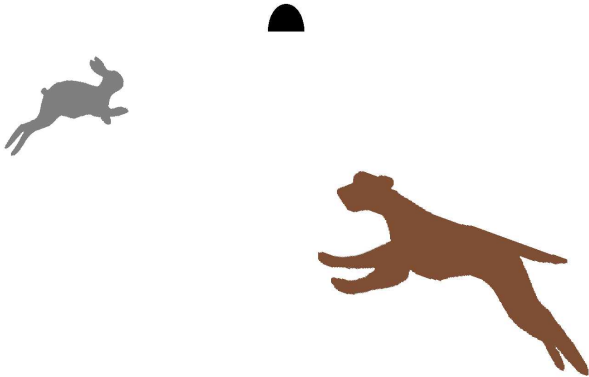
The former may represent the locations and the latter the velocities. Both can be seen as parametric curves. Next is an illustration of a real-valued and a vector-valued 1-forms; i.e., $n = 2, m = 1$ and $n = 2, m = 2$ respectively:



The former may represent a flow of water along a system of pipes and the latter the same flow with possible leaks. Both can be seen as vector fields. The algebra of vectors allows us to reproduce the definitions from Chapter 1 in the new, multi-dimensional, in output, context. We just assume that a space of inputs \mathbf{R} with a cell decomposition and a space of outputs \mathbf{R}^m are given.

5.2. The pursuit curves

In this chapter, we will often refer to systems of ODEs as *vector ODEs* or simply ODEs. Imagine a hound running after a rabbit. Let’s investigate their mutual dynamics.



We assume that the hound is always heading *at* the rabbit while the rabbit is heading for its hole and never changes directions.

We build a *discrete model*. This means that the time progresses in increments:

$$t_0 = 0, \; t_{n+1} = t_n + \Delta t.$$

The main assumption is:

- During the time interval $[t_n, t_n + \Delta t]$, the hound will be running toward the spot where the rabbit was at time t_n .

Suppose their locations are:

- (x, y) is the location of the rabbit, and
- (p, q) is the location of the hound.

They are represented by *parametric curves* defined at the nodes of the standard partition of the real line:

- $(x_n, y_n) = (x(t_n), y(t_n))$ is the location of the rabbit after n increments, and
- $(p_n, q_n) = (p(t_n), q(t_n))$ is the location of the hound after n increments.

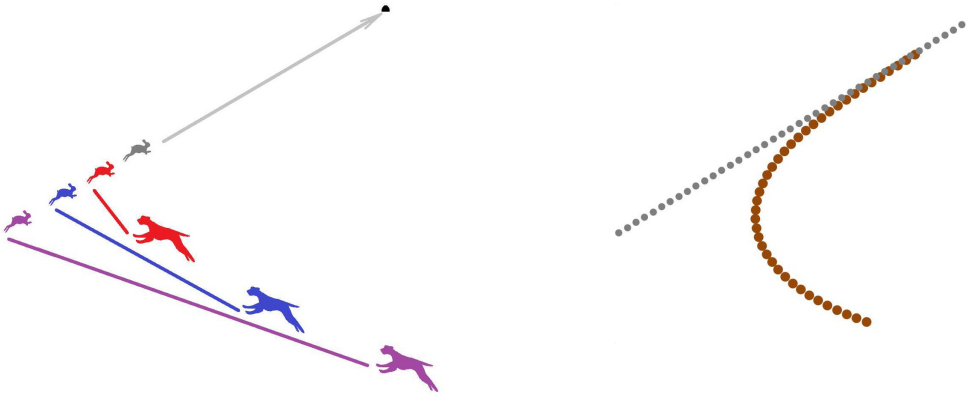
On the left we have, just as before, a simplified notation.

Suppose

- s is the speed of the rabbit, and
- v is the speed of the hound, $v > s$.

Example 5.2.1: independent of path

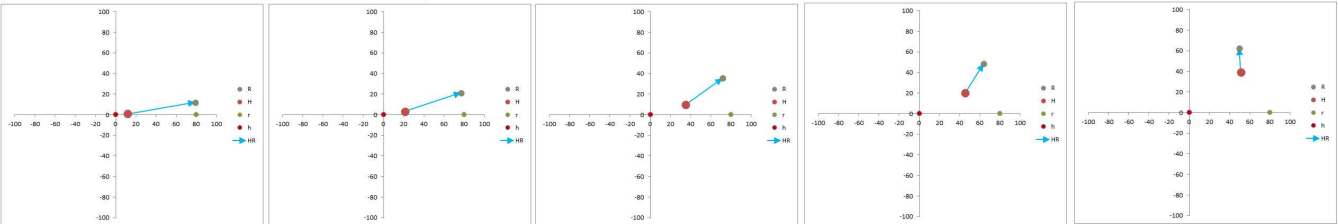
We create a model for the hound that is independent of a specific choice of the path of the rabbit: one step at a time.



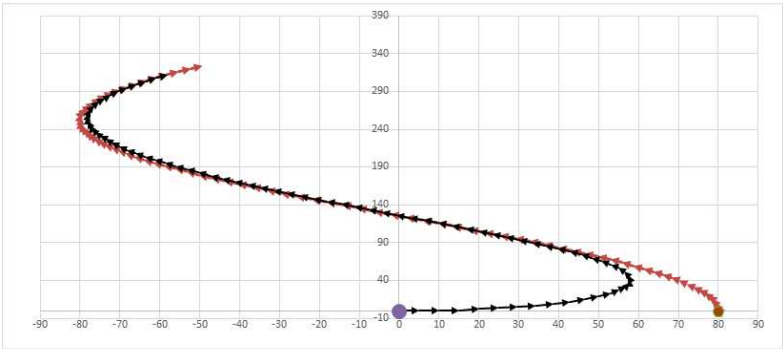
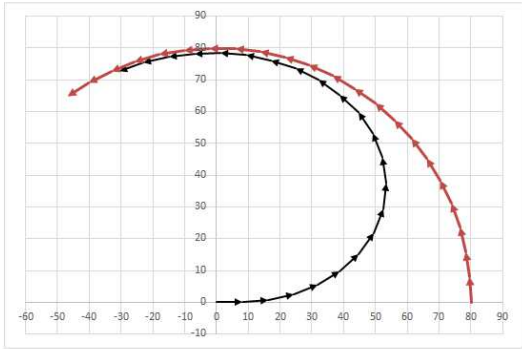
We compute consecutively:

- the difference between the two: $x_n - p_n$ and $y_n - q_n$, which is the direction from the hound towards the rabbit;
- the unit vector in this direction (dividing by the distance between them);
- the velocity of the hound (multiplying this vector by the hounds' speed v); and
- the new location (p_{n+1}, q_{n+1}) of the hound (adding the velocity times time interval $h = \Delta t$ to the last location).

	Rabbit	Hound							
t	speed	speed	direction	distance	unit direction	velocity of hound			
initial	0.00	80.00	0.00	80.00	0.00	80.00	1.00	0.00	80.00
last	0.14	79.22	11.16	11.16	0.77	68.05	10.40	68.04	0.99
current	0.15	79.00	11.90	11.90	0.80	67.15	11.07	66.05	0.99

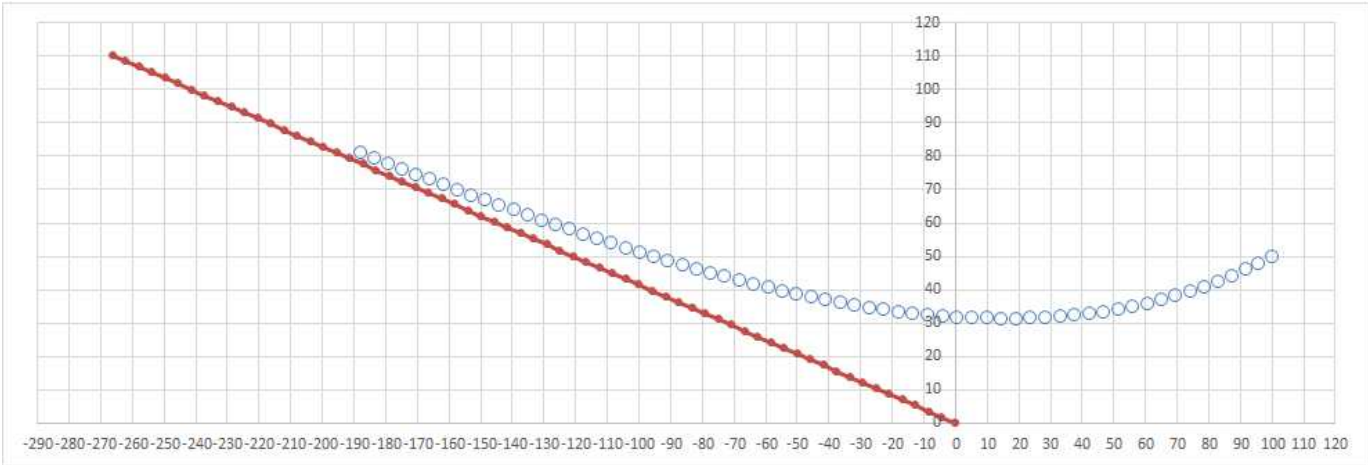


Above and below left show the pursuit of a rabbit following a circular path. Below right, it follows a sinusoid:



Exercise 5.2.2

Implement this model for the case of the rabbit running north-west:



Exercise 5.2.3

Implement this model for the case of the rabbit following a circle, sinusoid, spiral, etc.

We now approach this analytically. To make this possible, we choose a simple case: the rabbit is running along the y -axis from the origin. Then,

$$\begin{cases} x_{n+1} = 0, \\ y_{n+1} = y_n + s\Delta t. \end{cases}$$

Furthermore,

- $\Delta x_n = \Delta x(c_n)$
- $\Delta y_n = \Delta y(c_n)$
- $\Delta p_n = \Delta p(c_n)$
- $\Delta q_n = \Delta q(c_n)$

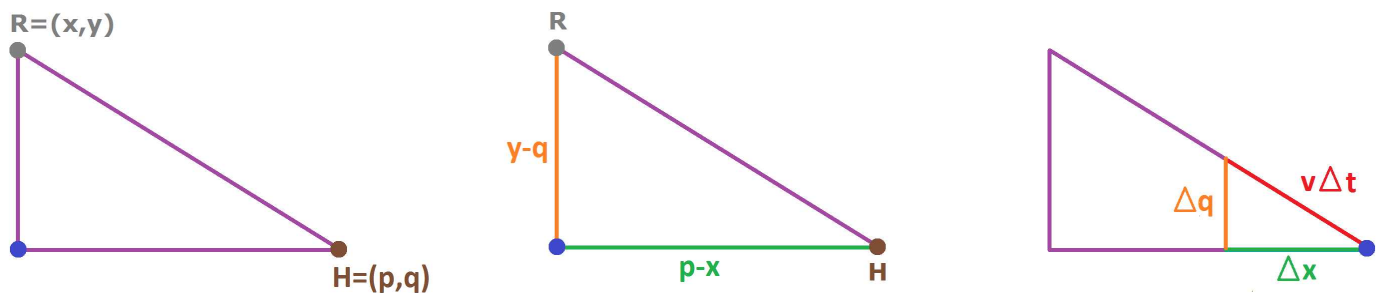
Here c_n are, say, the mid-points of the intervals of the partition.

For the hound, we have:

$$\begin{cases} p_{n+1} = p_n + \Delta p_n, \\ q_{n+1} = q_n + \Delta q_n. \end{cases}$$

We assume that the hound is located to the right of the y -axis and, therefore, moving to its left, $\Delta x_n < 0$. As the hound moves from (p_n, q_n) in the direction of (x_n, y_n) , the two right triangles are similar:

- first, the one with the sides $p_n - x_n$ and $y_n - q_n$, and
- second, the one with the sides $-\Delta p_n$ and Δq_n .



These observations give us the discrete model for the hound. We have for the tangent of the base angle of these triangles:

$$\tau_n = \frac{y_n - q_n}{p_n - x_n} = \frac{\Delta q_n}{-\Delta p_n}.$$

Therefore, we have a recursive formula for Δq_n ,

$$\Delta q_n = -\tau_n \cdot \Delta p_n,$$

if we can just find Δp_n . The hypotenuse of the latter triangle is

$$(\Delta p_n)^2 + (\Delta q_n)^2 = (v \cdot \Delta t)^2.$$

We derive from the last equation the following:

$$1 + \tau_n^2 = 1 + \left(\frac{\Delta q_n}{\Delta p_n}\right)^2 = \left(\frac{v \cdot \Delta t}{\Delta p_n}\right)^2.$$

Solving this equation for Δp_n , we choose the negative sign for the square root:

$$\Delta p_n = -\frac{v}{\sqrt{1 + \tau_n^2}} \Delta t.$$

Warning!

If the rabbit gets to the right of the hound, one has to change the sign.

Example 5.2.4: spreadsheet

Let's confirm our analysis with a spreadsheet. The setting are the following:

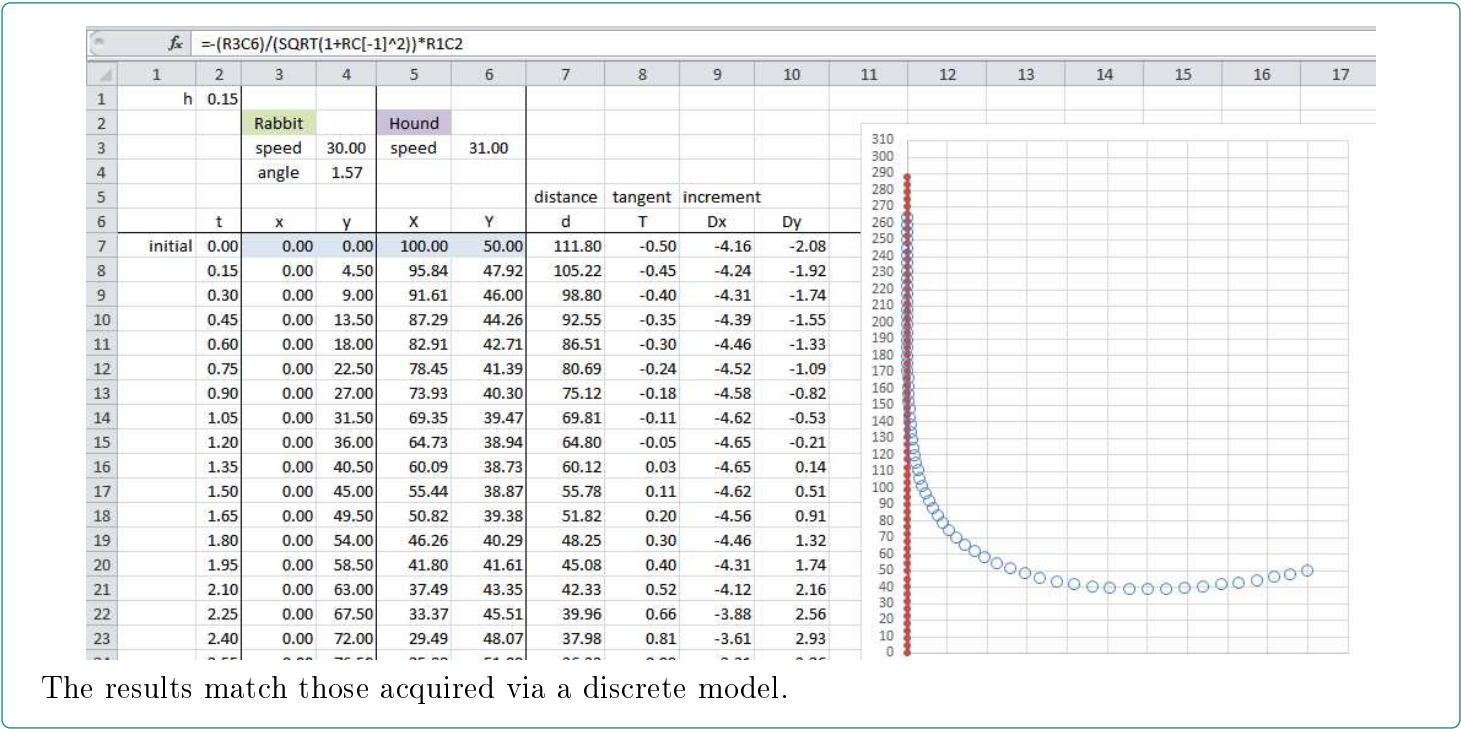
- $\Delta t = 0.1$ seconds
- $s = 30$ feet per second
- $v = 31$ feet per second
- $(x_0, y_0) = (0, 0)$ feet
- $(p_0, q_0) = (100, 50)$ feet

The main formulas are the ones for Δp_n and Δq_n :

$$=-(R3C6)/(SQRT(1+RC[-1]^2))*R1C2$$

$$=-RC[-2]*RC[-1]$$

These is the simulation:



From the two recursive formulas, we derive via substitution:

$$\Delta p_n = -\frac{v}{\sqrt{1+\tau_n^2}}\Delta t \implies \frac{\Delta p_n}{\Delta t} = -\frac{v}{\sqrt{1+\tau_n^2}}$$
$$\Delta q_n = -\tau_n \cdot \Delta p_n \implies \frac{\Delta q_n}{\Delta t} = \frac{v\tau_n}{\sqrt{1+\tau_n^2}}$$

Then, via $\Delta t \rightarrow 0$, we arrive at two time-dependent ODEs:

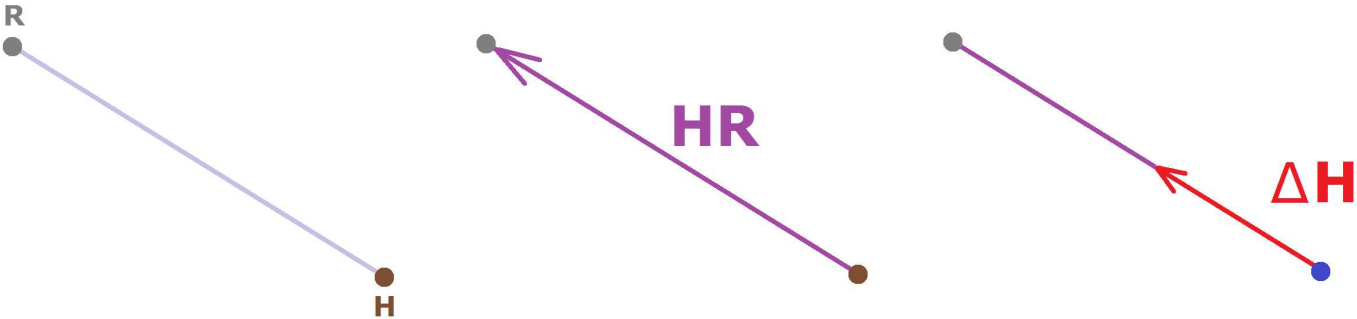
$$\begin{cases} \frac{dp}{dt} = -\frac{v}{\sqrt{1+\tau^2}}, \\ \frac{dq}{dt} = \frac{v\tau}{\sqrt{1+\tau^2}}, \end{cases}$$

where

$$\tau = \frac{y-q}{p-x},$$

and (x,y) is the *fixed* parametric curve representing the motion of the rabbit.

Now, let’s see what this analysis looks like in the *vector notation*. Suppose, as before, H is the location of the hound and R is the location of the rabbit. The latter is known. For the former, the displacement vector ΔH of H has to be collinear to the vector from H to R .



Therefore,

$$\Delta H \cdot HR = ||\Delta H|| \cdot ||HR|| = S||HR||.$$

The formula applies to any mutual location of the rabbit and the hound as well to pursuits in a space of any dimension.

Exercise 5.2.5

Derive an explicit ODEs for the case when the rabbit follows a straight path.

5.3. ODEs of second order as systems

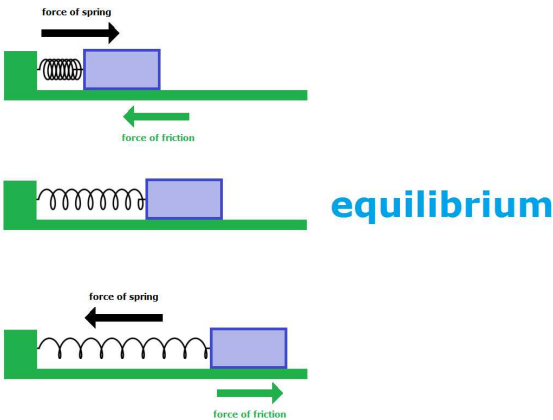
We previously discussed how ODEs of second order model motion of objects affected by forces. A special case is a *linear* ODE:

$$\frac{\Delta^2 x}{\Delta t^2} = a \frac{\Delta x}{\Delta t} + bx \text{ and } x'' = ax' + bx.$$

Here, we have:

- x is the position.
- $\frac{\Delta x}{\Delta t}$ and x' is the velocity.
- $\frac{\Delta^2 x}{\Delta t^2}$ and x'' is the acceleration.

Meanwhile a and b are constant numbers that express the proportional dependence of the force (i.e., the acceleration) on the velocity and the position respectively. For example, the object may be attached to the wall by a spring and, at the same time, be affected by the friction of the surface.



In fact, it makes sense to assume the following:

- The spring has its Hooke’s force proportional to the displacement from the equilibrium, with coefficient b , and since it pulls in the direction opposite to the displacement, we conclude that $b \leq 0$; and
- The friction force is proportional to the velocity, with coefficient a , and since it pulls in the direction opposite to the velocity, we conclude that $a \leq 0$.

These will be our assumptions:

$$a \leq 0, \quad b < 0.$$

Previously, we used the discrete equation above to the motion of spring and other dynamics. What about the other equation? In the simplest case $a = 0$, it was also solved. In our attempt to solve this equation in the general case we realize that we have no methods developed for equations of second order! When we encounter a completely new situation, we should always try to reduce it to something familiar.

We will apply the following *clever trick*. We introduce an extra (dependent) variable:

$$y = x'.$$

In terms of motion, the trick is nothing but re-introducing the velocity back into the picture. The result is a system of two ODEs:

$$\begin{cases} x' &= & y, \\ y' &= bx + ay. \end{cases}$$

The result is a trade-off: increasing the dimension – from 1 to 2 – but decreasing the order – from 2 for 1 – of the system. This 2×2 linear vector ODE,

$$X' = FX, \text{ with } F = \begin{bmatrix} 0 & 1 \\ b & a \end{bmatrix},$$

can now be subjected to the classification analysis presented in the last chapter.

First, the characteristic polynomial is:

$$\chi(\lambda) = \lambda^2 - a\lambda - b.$$

Suppose λ_1 and λ_2 are its two roots:

$$\lambda_{1,2} = \frac{a}{2} \pm \frac{\sqrt{a^2 + 4b}}{2}.$$

Then the outcomes depend on the sign of the discriminant,

$$D = a^2 + 4b.$$

Consider the case of $D \geq 0$. Since $a \leq 0$, we have

$$\lambda_1 = \frac{a}{2} - \frac{\sqrt{a^2 + 4b}}{2} < 0.$$

Since $b < 0$, we conclude that

$$\lambda_1 = \frac{a}{2} + \frac{\sqrt{a^2 + 4b}}{2} < \frac{a}{2} + \frac{\sqrt{a^2}}{2} = \frac{a}{2} + \frac{-a}{2} = 0$$

Therefore, an important conclusion is that *the real eigenvalues are always negative*.

According to our classification theorem, there are three main cases:

- Case 1: two distinct real roots;
- Case 2: one real repeated root;
- Case 3: complex conjugate roots.

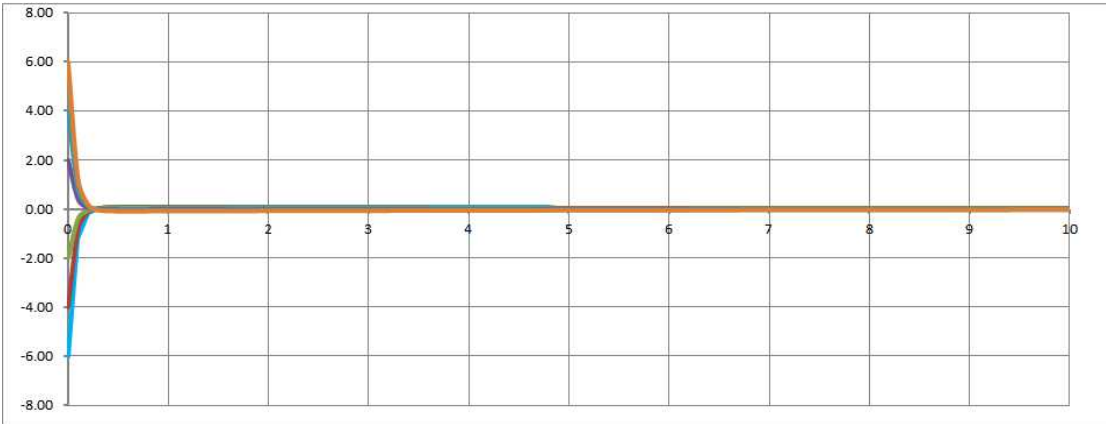
Case 1 is the case of

$$D = a^2 + 4b > 0.$$

The general solution is given by

$$x = Ce^{\lambda_1 t} + Ke^{\lambda_2 t}.$$

As $\lambda_1, \lambda_2 < 0$, we have exponential decay (a stable node of the vector ODE): the friction brings the ODE back to the equilibrium without oscillating. Here Euler’s method is run with $a = -8$, $b = -1$:



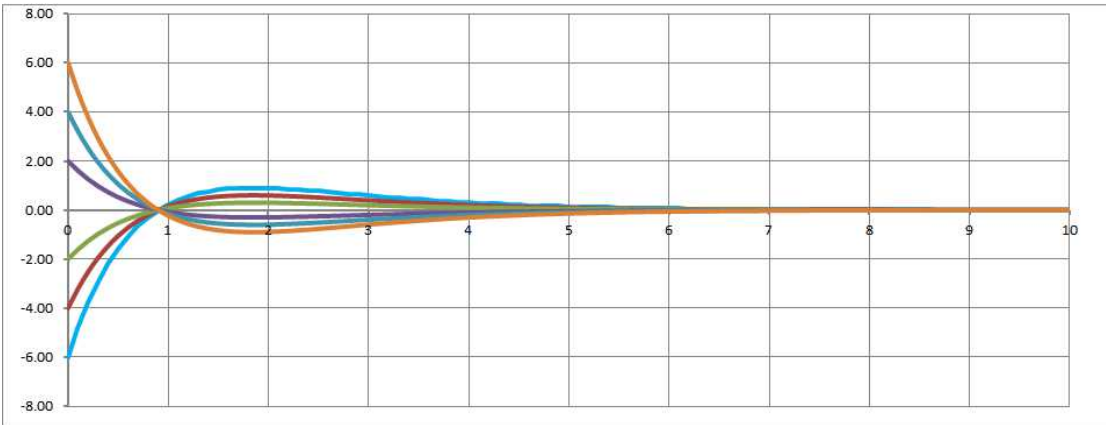
Case 2 is the case of

$$D = a^2 + 4b = 0 .$$

Again, as $\lambda = \lambda_1 = \lambda_2 < 0$, we see exponential decay (a stable improper node of the ODE): the ODE’s motion dies out, even though it might overshoot. The general solution is given by

$$x = (C + Kt)e^{\lambda t} .$$

Here Euler’s method is run with $a = -2$, $b = -1$:



Case 3 is the case of

$$D = a^2 + 4b < 0 .$$

The general solution is given by

$$x = Ce^{\alpha t} \cos(\beta t) + Ke^{\alpha t} \sin(\beta t) ,$$

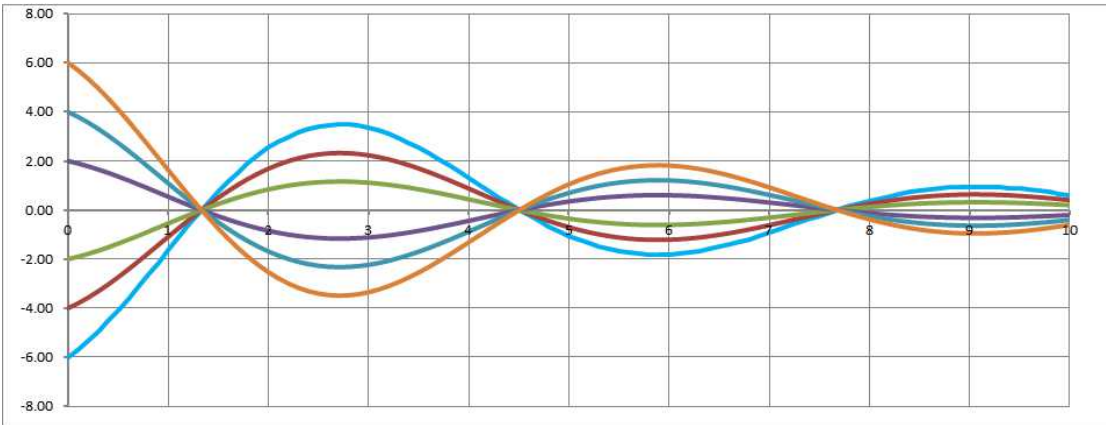
where

$$\alpha = \operatorname{Re} \lambda_{1,2} = \frac{a}{2}$$

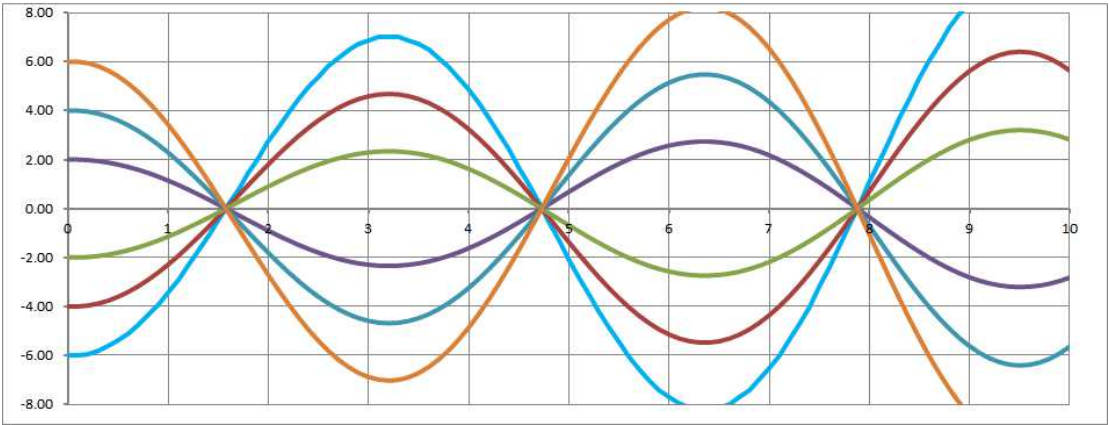
is non-negative by assumption and

$$\beta = \operatorname{Im} \lambda_{1,2} = \pm \frac{\sqrt{D}}{2}$$

is non-zero because $D \neq 0$. When $a < 0$, we see an oscillating decay (a stable focus of the ODE): the friction is strong enough to bring the ODE to the equilibrium eventually but isn’t strong enough to stop the ODE from oscillating. Here Euler’s method is run with $a = -.5$, $b = -1$:



When $a = 0$, we see pure oscillation (a center of the ODE): no friction. Here Euler’s method is run with $a = 0$, $b = -1$:



Looking at these illustration as a sequence reveals the effect of growing of the friction coefficient b .

Exercise 5.3.1

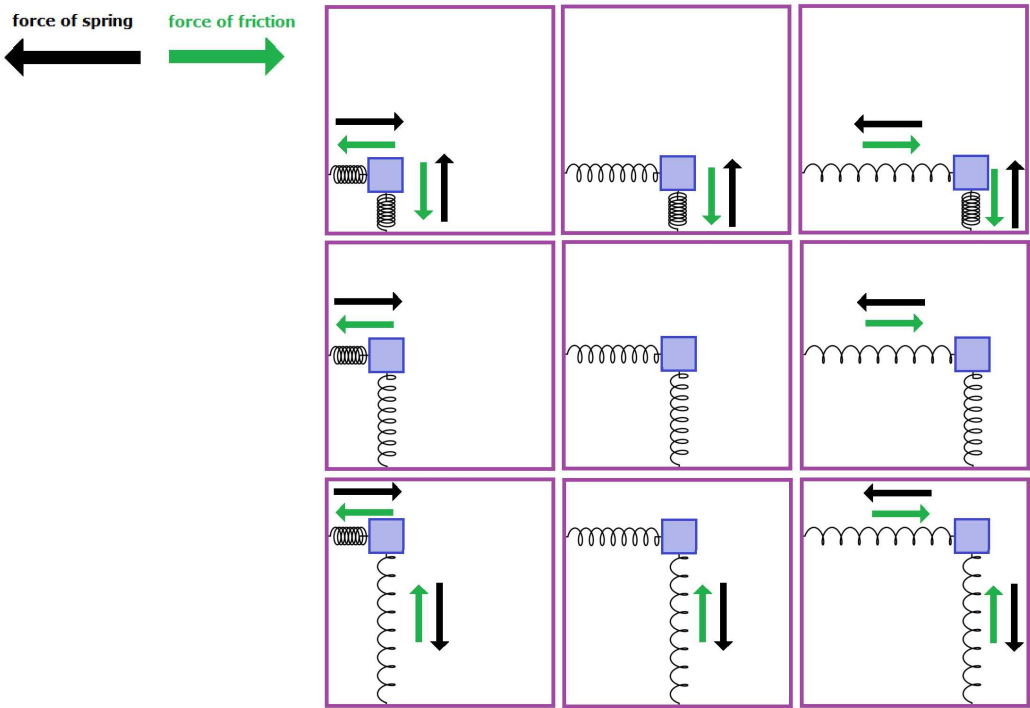
Point out in what way the last illustration doesn't match the description and explain why.

5.4. Vector ODEs of second order: a double spring

What of the nature of the forces remains the same but the motion is *on the plane*?

In the simplest situation, we have this:

- Two different springs are attached to the object along the two axes.
- Two different kinds of friction are produced along the two axes.



As the object moves, we have two effects:

- Two springs exert two forces along the axes – negatively proportional to the displacement from the equilibrium.
- Two kinds of friction exert two forces along the axes – negatively proportional to the velocity.

The displacements and the velocities of the object in the two directions are independent of each other. Therefore, there will be *two* linear ODE of order 2 of the same kinds as before:

$$x'' = ax' + bx \quad \text{and} \quad y'' = cy' + dy,$$

where

$$a \leq 0 \quad \text{and} \quad b \leq 0 \quad \text{and} \quad c \leq 0 \quad \text{and} \quad d \leq 0.$$

In this case, the classification in the last section applies, separately. We combine the two *scalar* equations into one *vector* equation:

$$\begin{cases} x'' = ax' + bx \\ y'' = cy' + dy \end{cases} \quad \text{leading to} \quad \begin{bmatrix} x'' \\ y'' \end{bmatrix} = \begin{bmatrix} a & 0 \\ 0 & c \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix} + \begin{bmatrix} b & 0 \\ 0 & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Generally, we have a vector ODE:

$$X'' = AX' + BX.$$

Here,

$$X = \begin{bmatrix} x \\ y \end{bmatrix}, \quad X' = \begin{bmatrix} x' \\ y' \end{bmatrix}, \quad X'' = \begin{bmatrix} x'' \\ y'' \end{bmatrix}$$

are the position vector, the velocity, and the acceleration, respectively. Also, A and B are two 2×2 matrices that express the (linear) dependence of the force (i.e., the acceleration) on the velocity and position respectively. They don't have to be diagonal anymore.

In contrast to the last section, we will concentrate on the *discrete* ODEs:

$$\frac{\Delta^2 X}{\Delta t^2} = A \frac{\Delta X}{\Delta t} + BX.$$

Here, a solution X is a parametric curve defined on the nodes of a partition of an interval, t_0, t_1, \dots (i.e., a discrete 0-form), so that its difference quotient and the second difference quotient satisfy the equation:

$$X = \begin{bmatrix} x \\ y \end{bmatrix}, \quad \frac{\Delta X}{\Delta t} = \begin{bmatrix} \frac{\Delta x}{\Delta t} \\ \frac{\Delta y}{\Delta t} \end{bmatrix}, \quad \frac{\Delta^2 X}{\Delta t^2} = \begin{bmatrix} \frac{\Delta^2 x}{\Delta t^2} \\ \frac{\Delta^2 y}{\Delta t^2} \end{bmatrix}.$$

In order to simplify the notation, we define three sequences of vectors in \mathbf{R}^2 :

$$\begin{aligned} X''_n &= \frac{\Delta^2 X}{\Delta t^2}(t_n) \\ X'_n &= \frac{\Delta X}{\Delta t}(t_n) \\ X_n &= X(t_n) \end{aligned}$$

The ODE then produces these three recursive formulas for those three sequences of vectors:

$$\begin{aligned} X''_{n+1} &= AX'_n + BX_n \\ X'_{n+1} &= X'_n + X''_{n+1} \Delta t \\ X_{n+1} &= X_n + X'_{n+1} \Delta t \end{aligned}$$

We apply them below.

The starting point is the simplest case of the two springs. Here, the matrices are *diagonal*:

$$A = \begin{bmatrix} a_x & 0 \\ 0 & a_y \end{bmatrix}, \quad B = \begin{bmatrix} b_x & 0 \\ 0 & b_y \end{bmatrix},$$

with

$$a_x, a_y \leq 0 \quad \text{and} \quad b_x, b_y < 0.$$

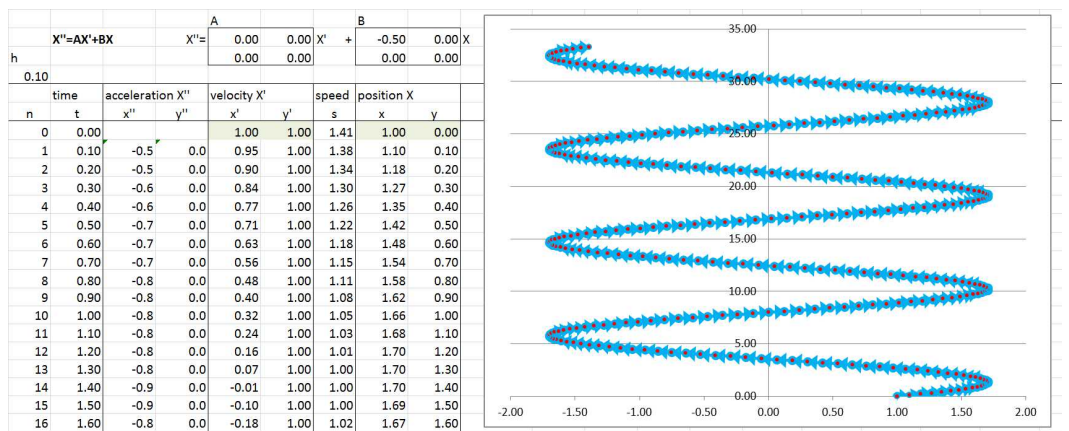
Below, we will be gradually adding new forces to the setup by adding a new non-zero entry to the matrices. The results are illustrated with the corresponding discrete models produced by Euler’s method.

The initial conditions will be the same:

$$X_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad X'_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

We start with nothing but a *spring* along the x -axis:

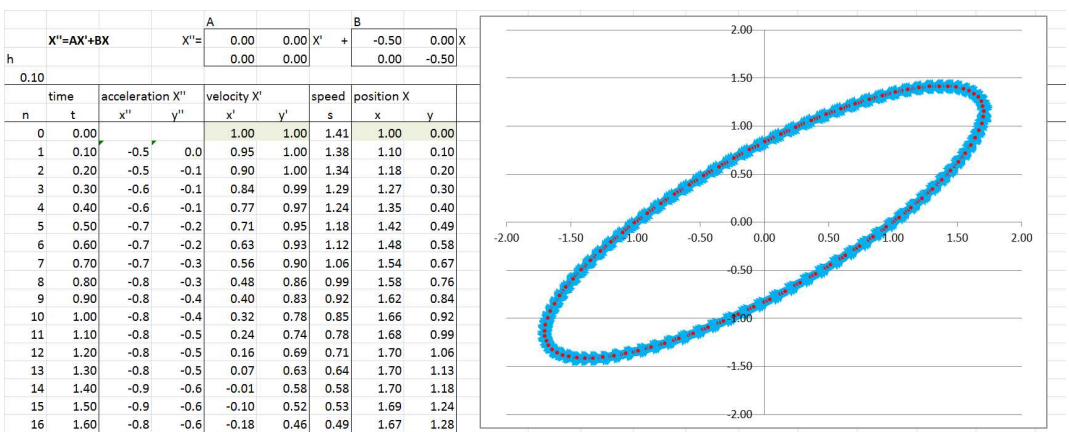
$$A = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} -0.5 & 0 \\ 0 & 0 \end{bmatrix}.$$



The result is an oscillation along the x -axis and uniform motion along the y -axis.

We now add an *identical spring* for the y -axis:

$$A = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} -0.5 & 0 \\ 0 & -0.5 \end{bmatrix}.$$



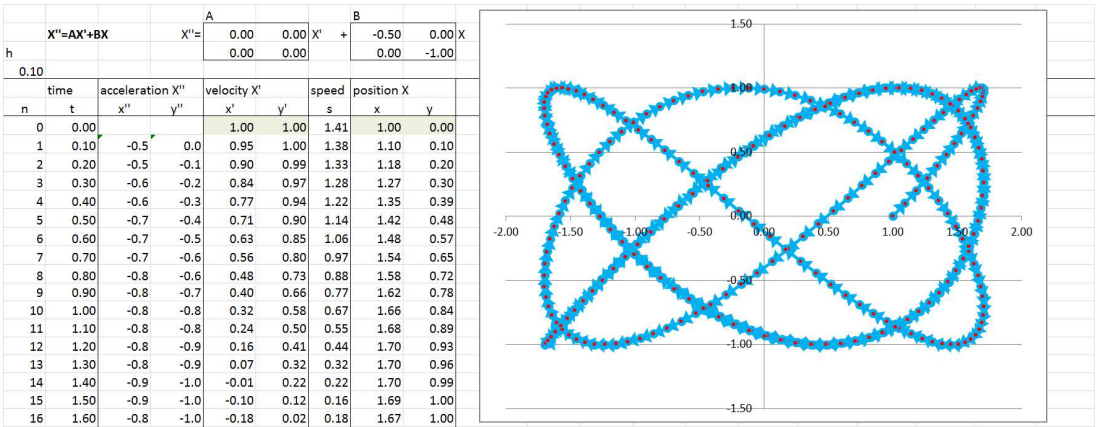
As a result, both x and y oscillate at the same period and come back simultaneously; it’s a cycle. The path appears to be an ellipse centered at the origin. We also notice that the motion is slows down at the tips of this ellipse.

Exercise 5.4.1

Why isn’t the trajectory a circle?

Let’s make the latter *spring more rigid*:

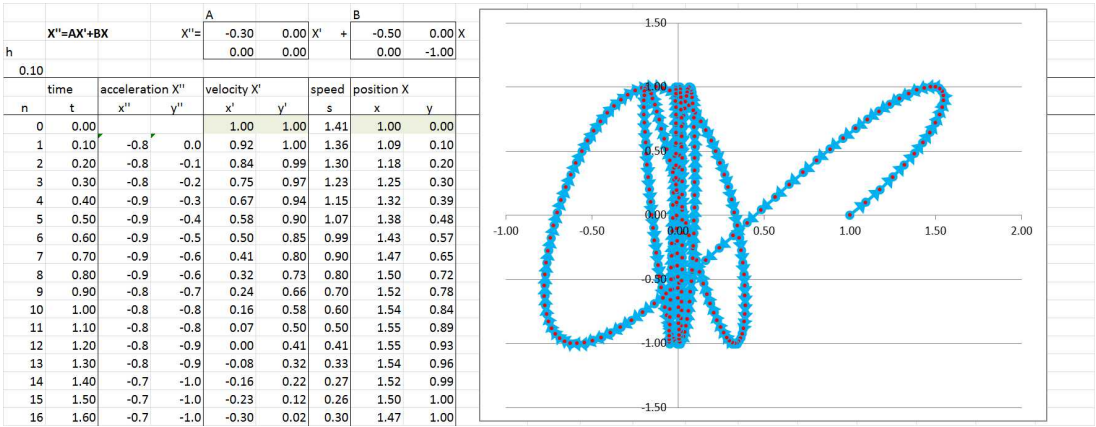
$$A = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} -0.5 & 0 \\ 0 & -1 \end{bmatrix}.$$



As a result, x and y still oscillate but at different frequencies. They may come back simultaneously after several periods. That will create a cycle.

We now add *friction* in the x -direction:

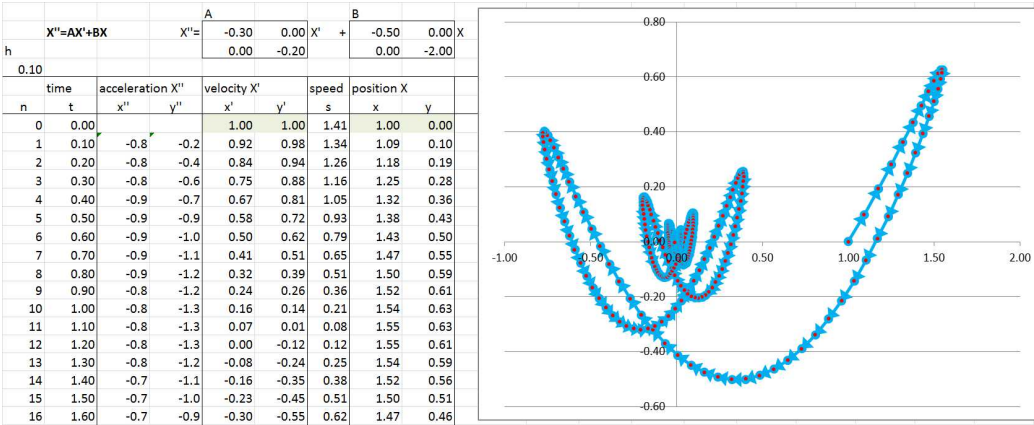
$$A = \begin{bmatrix} -0.3 & 0 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} -0.5 & 0 \\ 0 & -1. \end{bmatrix}.$$



While y still oscillates, the x -position's oscillation diminishes as it approaches 0.

We also add *friction* in the y -direction now:

$$A = \begin{bmatrix} -0.3 & 0 \\ 0 & -0.2 \end{bmatrix}, \quad B = \begin{bmatrix} -0.5 & 0 \\ 0 & -1.0 \end{bmatrix}.$$

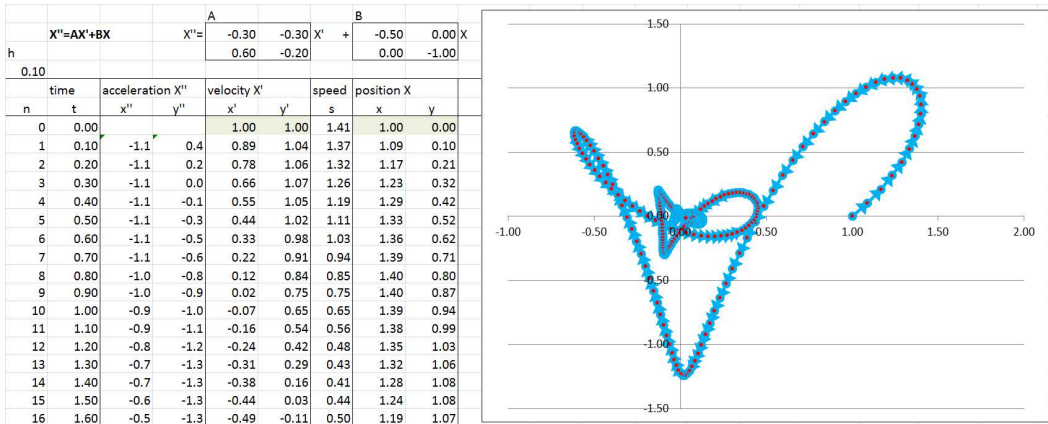


Both x or y oscillations diminish and the point approaches 0.

Up until now, x and y has been independent and the two ODEs are solvable separately. This is why everything we see in the simulations are confirmed by the classification in the last section. From now on, this classification doesn't apply anymore. Note also that the clever trick of introducing the velocities as new variables will produce 4×4 matrices, the analysis of which lies outside the scope of this text. This is why we proceed with discrete models only.

We intermix the two variables now by making the *friction vary* in the directions other than the two axes (for example, the springs may be placed under various angles):

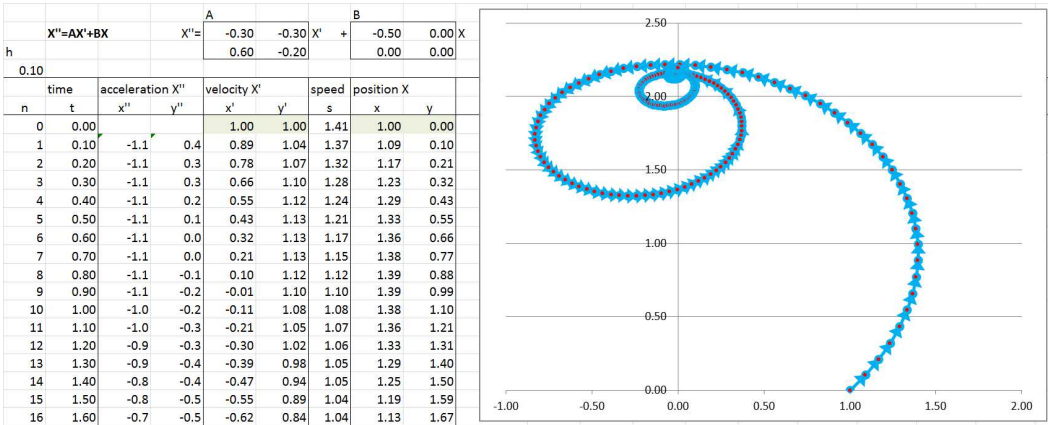
$$A = \begin{bmatrix} -0.3 & -\mathbf{0.3} \\ \mathbf{0.6} & -0.2 \end{bmatrix}, \quad B = \begin{bmatrix} -0.5 & 0 \\ 0 & -1 \end{bmatrix}.$$



The pattern of back-and-forth convergence toward 0 remains.

As a final experiment, we *remove* the y -axis spring:

$$A = \begin{bmatrix} -0.3 & -0.3 \\ 0.6 & -0.2 \end{bmatrix}, \quad B = \begin{bmatrix} -0.5 & 0 \\ 0 & \mathbf{0} \end{bmatrix}.$$



The oscillation of the y -coordinate moves away from 0.

Let’s look at what happens to the *energy* in this discrete model. But first let’s see what the data tells about the continuous motion.

For simplicity we continue to assume that the mass is equal to 1. Then the *kinetic energy* is known to be this function of time:

$$K(t) = \frac{1}{2} ||X'(t)||^2.$$

Euler’s method samples the velocity and gives us an approximation of this function, the *sampled kinetic energy*:

$$K_n = \frac{1}{2} ||X'_n||^2,$$

in each row of the spreadsheet. Next, suppose that *the force is conservative*. Then the *potential energy* can be found as the work of this force:

$$W(t) = - \int_0^t X'' \cdot X' dt.$$

What does the data tell us about this function? Once again, Euler’s method samples the velocity and gives us an approximation of this function, the *sampled potential energy* computed as follows. Over a period of time from t_{k-1} to t_k (length Δt), the potential energy grows by:

$$W_k = -X''_k \cdot X'_k \Delta t.$$

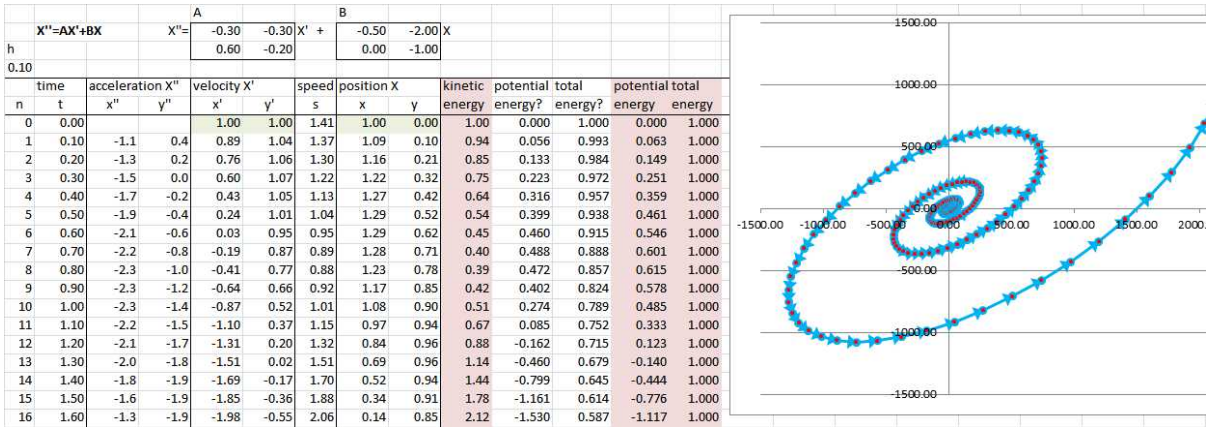
We add these in all rows up to the n th to find the current *sampled potential energy*:

$$P_n = \sum_{k=0}^n W_k.$$

The next column contains the sum of the two, the *total sampled energy*:

$$E_n = K_n + P_n = \frac{1}{2}||X'_n||^2 - \sum_{k=0}^n X''_k \cdot X'_k \Delta t.$$

As you can see, it's not conserved!



The result is to be expected from sampling (a discrete approximation of) a continuous motion.

Definition 5.4.2: discrete ODE of second order

A *discrete ODE of second order* is given by the following recursive formulas:

$$X''_{n+1} = f(X_n, X'_n)$$
$$X'_{n+1} = X'_n + X''_{n+1} \Delta t$$
$$X_{n+1} = X_n + X'_{n+1} \Delta t$$

for some function f and a non-zero number Δt . Its *solution* is three sequences of vectors that satisfies these equations.

The discrete ODE conserves energy if we just look at the work the right way. We recompute the potential energy in the next column with the current velocity X'_k replaced with the average of the current and the previous velocities:

$$W_k = X''_k \cdot \frac{1}{2}(X'_{k-1} + X'_k) \Delta t.$$

The last column confirms the conservation of energy. This fact is also easy to prove algebraically for all linear or non-linear discrete systems.

Theorem 5.4.3: Conservation of Energy

The *energy* of a solution X of a discrete ODE of second order defined as

$$E_n = \frac{1}{2}||X'_n||^2 - \sum_{k=0}^n X''_k \cdot \frac{1}{2}(X'_{k-1} + X'_k) \Delta t,$$

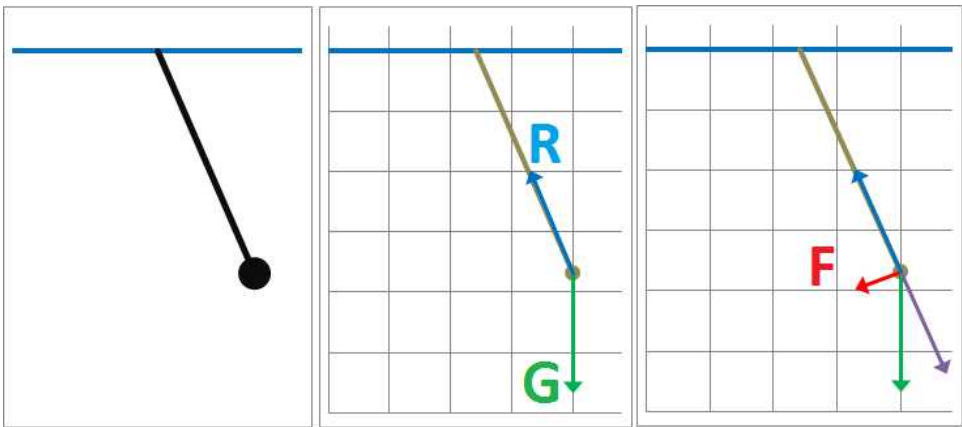
is constant.

Proof.

$$\begin{aligned} E_{n+1} - E_n &= \frac{1}{2} \|X'_{n+1}\|^2 - \frac{1}{2} \|X'_n\|^2 - X''_{n+1} \cdot \frac{1}{2} (X'_n + X'_{n+1}) \Delta t \\ &= \frac{1}{2} (\|X'_{n+1}\|^2 - \|X'_n\|^2) - \frac{1}{\Delta t} (X'_{n+1} - X'_n) \cdot \frac{1}{2} (X'_{n+1} + X'_n) \Delta t \\ &= \frac{1}{2} (\|X'_{n+1}\|^2 - \|X'_n\|^2) - \frac{1}{2} (X'_{n+1} \cdot X'_{n+1} - X'_n \cdot X'_n) \\ &= 0. \end{aligned}$$

5.5. A pendulum

Description: “an object is attached to a fixed point with a string and subjected to the gravity”.
Let x be the horizontal axis and z the vertical.



We compute what’s necessary, i.e., the acceleration, for the spreadsheet discussed above. Suppose the length of the string is L . First, the *gravity* is the constant vector

$$G = \langle 0, -gm \rangle .$$

Then the *resistance* R of the string is found as the negative of the projection of the gravity on the line from 0 to the current location $X = \langle x, z \rangle$:

$$R = -\frac{G \cdot X}{\|X\|^2} X = -\frac{G \cdot X}{L^2} X .$$

Then we find the *tangential component of the gravity*:

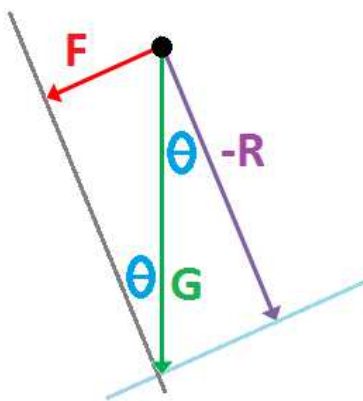
$$F = R + G .$$

Or, it is simply:

$$F = G \sin \theta ,$$

where θ is the angle of the string with the vertical. Note that the horizontal component of the force is zero when the string is vertical and when it is horizontal and it reverses its direction. Meanwhile, the vertical component is always negative. And the tangential acceleration is:

$$A = F/m .$$



These quantities are computed in that order just as before. The updated values of the velocities and the locations are also found except z is found from x by:

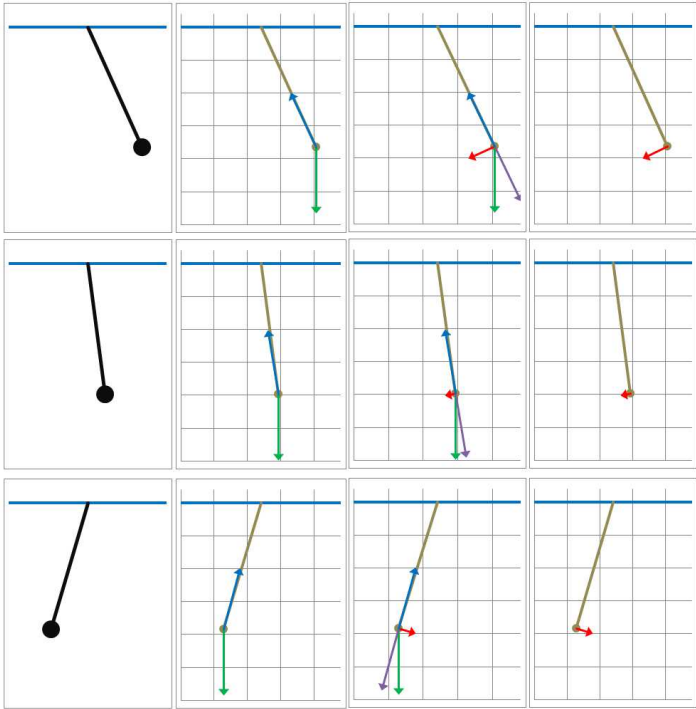
$$x^2 + y^2 = L^2,$$

where L is the length of the string. Equivalently,

$$x = L \cos \theta, \quad y = L \sin \theta.$$

We use a spreadsheet to evaluate these column by column and then for each iteration of t :

h= 0.10													
	time		gravity		dot	decomposition G=R+F				acceleration F/m/velocity			
	n	t	m	G		R	F			x''	z''	x'	z'
	0	0.00	1.00	0.00	-1.00	1.81	0.38	-0.82	-0.38	-0.18		0.00	0.00
	1	0.10	1.00	0.00	-1.00	1.81	0.38	-0.82	-0.38	-0.18		-0.04	-0.02
t-2	1	0.10	1.00	0.00	-1.00	1.81	0.38	-0.82	-0.38	-0.18		-0.04	-0.02
t-1	3	0.20	1.00	0.00	-1.00	1.81	0.38	-0.82	-0.38	-0.18		-0.08	-0.04
t	4	0.30	1.00	0.00	-1.00	1.82	0.38	-0.82	-0.38	-0.18		-0.11	-0.05

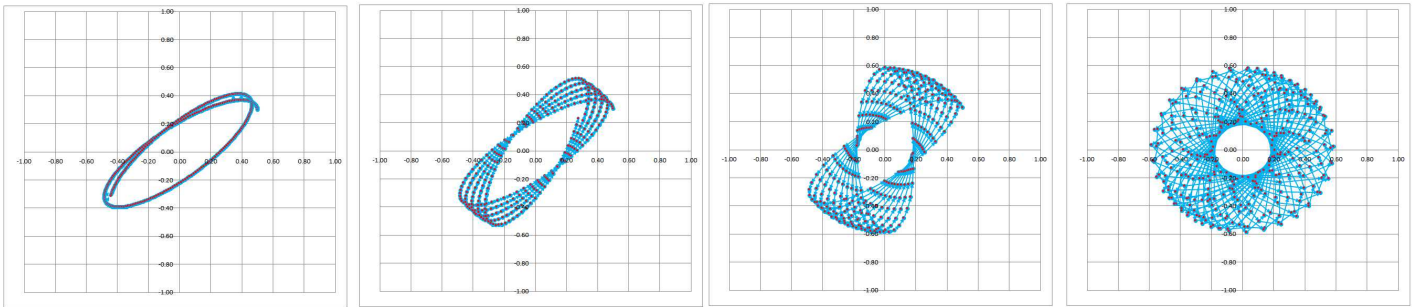


The visualization suggests that we indeed have a pendulum.

For the 3-dimensional case, we give the pendulum another – horizontal – degree of freedom, y . The vector analysis remains the same with $G = \langle 0, 0, -gm \rangle$ and $X = \langle x, y, z \rangle$. The spreadsheet only requires adding columns for y . The values of z is found from:

$$x^2 + y^2 + z^2 = L^2.$$

Below we plot the parametric curve $\langle x, y \rangle$ with a non-zero horizontal component of the initial velocity:



We observe that the pendulum swings as before but the plane of the swings is continuously rotating. We now go back to the 2-dimensional case and address it analytically. We apply Newton’s law to the tangential axis only:

$$F = -mg \sin \theta = ma \implies a = -g \sin \theta .$$

How is this linear acceleration a along the tangent related to the change in angle θ ? Let s be the arc-length parameter. Since the curve is a circle of radius L , we have:

$$s = L\theta .$$

Now, the arc-length parameter is a function of t , i.e., $s = s(t)$. We differentiate twice with respect to t :

$$s = L\theta \implies v = s' = L\theta' \implies a = x'' = L\theta'' .$$

Therefore,

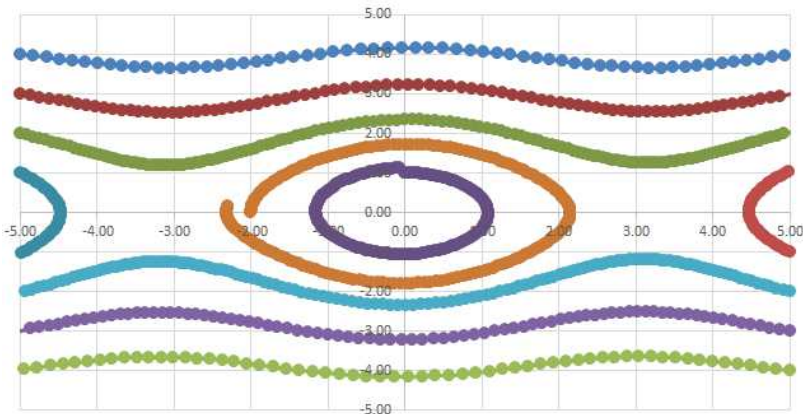
$$L\theta'' = -g \sin \theta ,$$

or

$$\theta'' + \frac{g}{L} \sin \theta = 0 .$$

This is a *non-linear* ODE with respect to θ .

To confirm our simulations, we plot a few solutions using Euler’s method just as in the last chapter:



Here θ is plotted against $\alpha = \theta'$; i.e., we are considering the following system:

$$\begin{cases} \theta' = \alpha, \\ \alpha' = -\frac{g}{L} \sin \theta. \end{cases}$$

The Euler’s solutions aren’t periodic but rather appear to be spirals!

Let’s *linearize*!

There is only one equilibrium $\theta = \alpha = 0$. Now,

$$\frac{d}{d\theta}(\sin \theta)(0) = 1 .$$

Therefore, the linearized systems is:

$$\begin{cases} \theta' = \alpha, \\ \alpha' = -\frac{g}{L}\theta. \end{cases}$$

We can look at the eigenvalues:

$$A = \begin{bmatrix} 0 & 1 \\ -\frac{g}{L} & 0 \end{bmatrix} \implies \chi_A(\lambda) = \det(A - \lambda I) = \det \begin{bmatrix} -\lambda & 1 \\ -\frac{g}{L} & -\lambda \end{bmatrix} = \lambda^2 + \frac{g}{L} = 0 \implies \lambda_{1,2} = \pm \sqrt{\frac{g}{L}}i.$$

According to the *Classification of Linear Systems II* in [Chapter 2](#), the system has a *center* at 0! Of course, we can just solve the linearized system:

$$\theta'' = \alpha' = -\frac{g}{L}\theta \implies \theta = \theta_0 \cos\left(\sqrt{\frac{g}{L}}t\right),$$

with the initial conditions:

$$\theta(0) = \theta_0 \text{ and } \theta'(0) = 0.$$

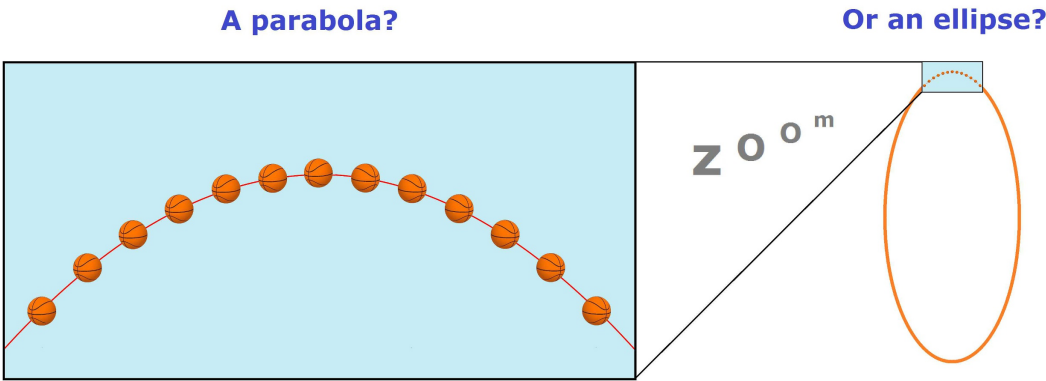
So, our conclusion is that when small the swings are close to periodic with the following period:

$$2\pi\sqrt{\frac{g}{L}}.$$

The period is independent of the amplitude θ_0 !

5.6. Planetary motion

A familiar problem about a ball thrown in the air has a solution: its trajectory is a *parabola*. However, we also know that if we throw really-really hard (like a rocket) the ball will start to orbit the Earth following an *ellipse*.



The motion of two planets (or a star and a planet, or a planet and a satellite, etc.) is governed by a single force: the *gravity*. Recall how this force operates.

Newton's Law of Gravity: The force of gravity between two objects is

- 1. proportional to either of their masses,
- 2. inversely proportional to the square of the distance between their centers, and
- 3. directed along the segment between them.

In other words, the force is given by the formula:

$$F = G \frac{mM}{r^2},$$

where:

- F is the force between the objects.
- G is the gravitational constant.
- m is the mass of the first object.
- M is the mass of the second object.
- r is the distance between the centers of the masses.

Let’s connect this familiar physics law to another.

Kepler’s Laws of Planetary Motion: The motion of planets around the Sun follows these laws:

1. The orbit of a planet is an ellipse with the Sun at one of the two foci.
2. A line segment joining a planet and the Sun sweeps out equal areas during equal intervals of time.
3. The square of the orbital period of a planet is proportional to the cube of the semi-major axis of its orbit.

We know the vector form of Newton’s law (with the first object located at the origin):

$$F = -GM \frac{X}{||X||^3}.$$

Then, we have a vector ODEs of second order for the location X of the second object:

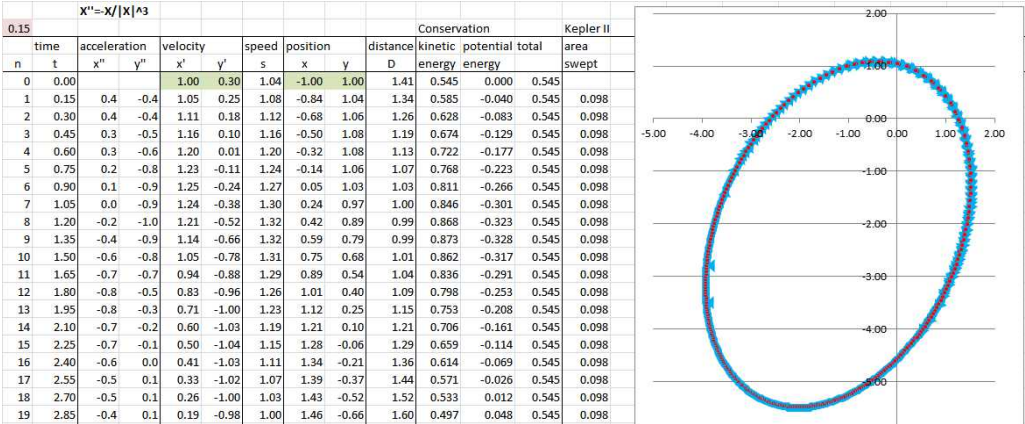
$$X'' = -GM \frac{X}{||X||^3}.$$

We will call it *Newton’s ODE of planetary motion*. It is non-linear.

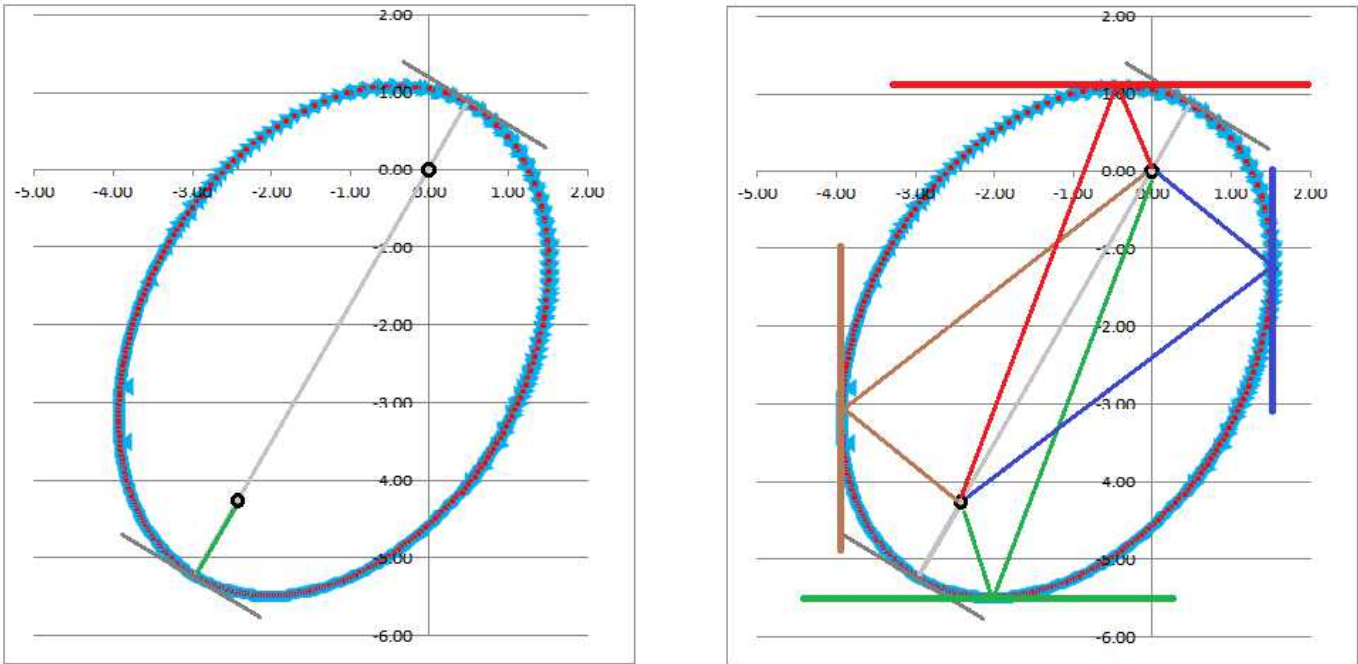
As before, from the acceleration X''_n the velocity X'_n and then from the velocity the position X_n are computed by what amounts to Euler’s method. The discrete model of planetary motion is a discrete model of second order given by the following recursive formulas:

$$\begin{aligned} X''_{n+1} &= -GM \frac{X_n}{||X_n||^3} \\ X'_{n+1} &= X'_n + X''_{n+1} \Delta t \\ X_{n+1} &= X_n + X'_{n+1} \Delta t \end{aligned}$$

The result is as predicted: the points form what looks like an ellipse with one of the foci at the origin (First Kepler’s Law).



In order to confirm this idea, we first draw a line from the origin (the first focus) to the point on the curve where it is perpendicular to the tangent. We then continue this line to the other end of the curve and discover that indeed the tangent is again perpendicular to this chord. We then measure the distance from the focus to the first point and plot the same distance from the second point along the chord. This gives us the second focus (left).

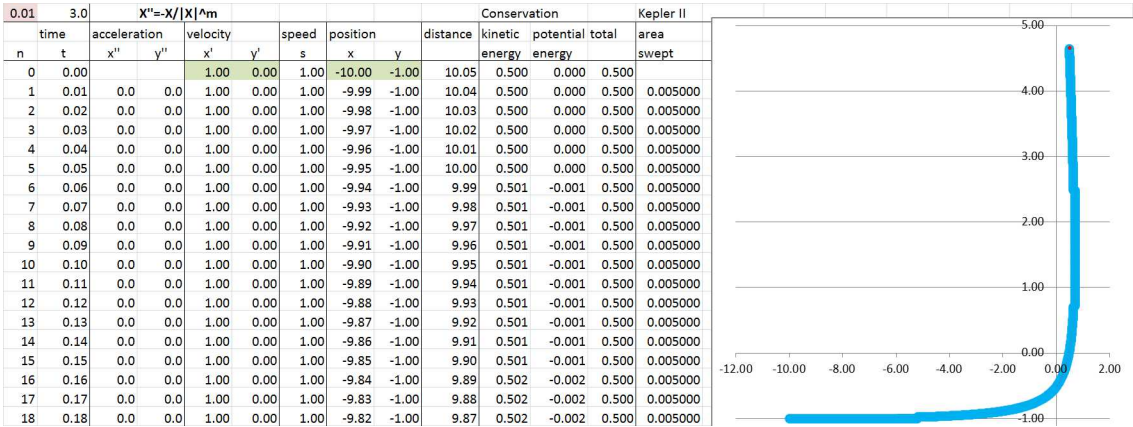


We further test our conjecture by plotting the lines representing rays of light that start at one focus and then bounce off the curve to the other focus (right).

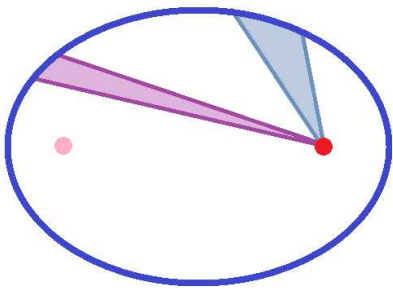
Exercise 5.6.1

Is this really an ellipse?

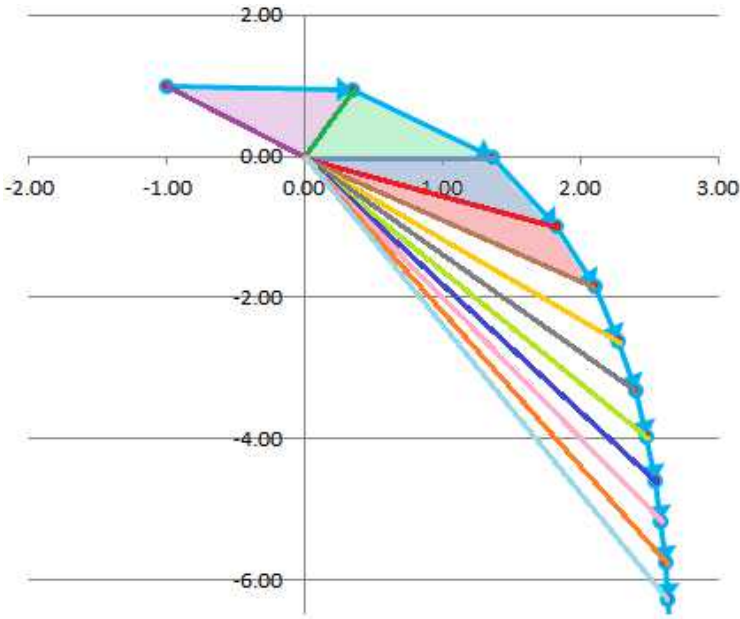
By changing the initial conditions we can produce other patterns of behavior such as what looks like a hyperbola:



Now what about the Second Kepler's Law?



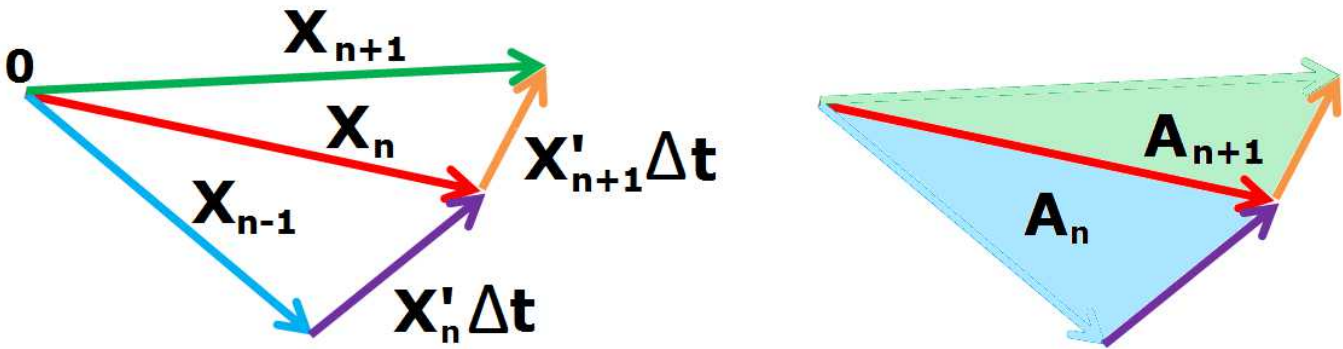
We can see in our simulation that the points away from the origin (the first object) are closer spaced than the ones closer to the origin. This means that the motion is faster when they are closer. That is why the longer triangles are thinner.



We confirm that the areas of the triangles are the same by computing them in the last column of the spreadsheet via the *Parallelogram Formula*:

$$A_{n+1} = \frac{1}{2} \left| \det [X_{n+1} X_n] \right|,$$

where X_{n+1} and X_n are given by their column vectors.



We will derive the law in a more general setting than the original. First we note that it might apply to *all* discrete trajectories not just ellipses.

Exercise 5.6.2

Prove that an object moving along a straight line at a constant speed satisfies the area property with respect to any fixed point. Hint: use elementary geometry.

Second, we will see that the law applies to more general forces than just the gravity. The gravity is a *central force*, i.e., one that is directed along the line between the location and the origin. Then the recursive formula for such a discrete models is:

$$X''_{n+1} = L(||X_n||)X_n ,$$

for some function L .

Theorem 5.6.3: Discrete Second Kepler’s Law

If X is a trajectory of the discrete model with central force, the segment from the focus to the location sweeps an equal area in an equal time. Conversely, if this area property is satisfied by each trajectory of a discrete model of order two, the model is produced by a central force.

Proof.

We consider the model:

$$X''_{n+1} = f(X_n) ,$$

where f is some function. Then,

$$X'_{n+1} = X'_n + X''_n \Delta t = X'_n + f(X_n) \Delta t .$$

We will use twice the Parallelogram Formula for the oriented area A of the triangle spanned by vectors M and N :

$$A = \frac{1}{2} \det [M \ N] .$$

However, the vectors chosen will be different from those used in the spreadsheet. First, suppose the triangle with area A_n is formed by X_n and $X'_n \Delta t$; therefore, we have by the linearity of the determinant:

$$A_n = \frac{\Delta t}{2} \det [X_n \ X'_n] .$$

Second, suppose the triangle with area A_{n+1} is formed by X_n and $X'_{n+1} \Delta t$; therefore, we have the following by the linearity and the additivity of the determinant:

$$\begin{aligned} A_{n+1} &= \frac{\Delta t}{2} \det [X_n \ X'_{n+1}] \\ &= \frac{\Delta t}{2} \det [X_n \ (X'_n + f(X_n) \Delta t)] \\ &= \frac{\Delta t}{2} \big(\det [X_n \ X'_n] + \Delta t \det [X_n \ f(X_n)] \big) . \end{aligned}$$

Therefore,

$$A_{n+1} - A_n = \frac{\Delta t^2}{2} \det [X_n \ f(X_n)] .$$

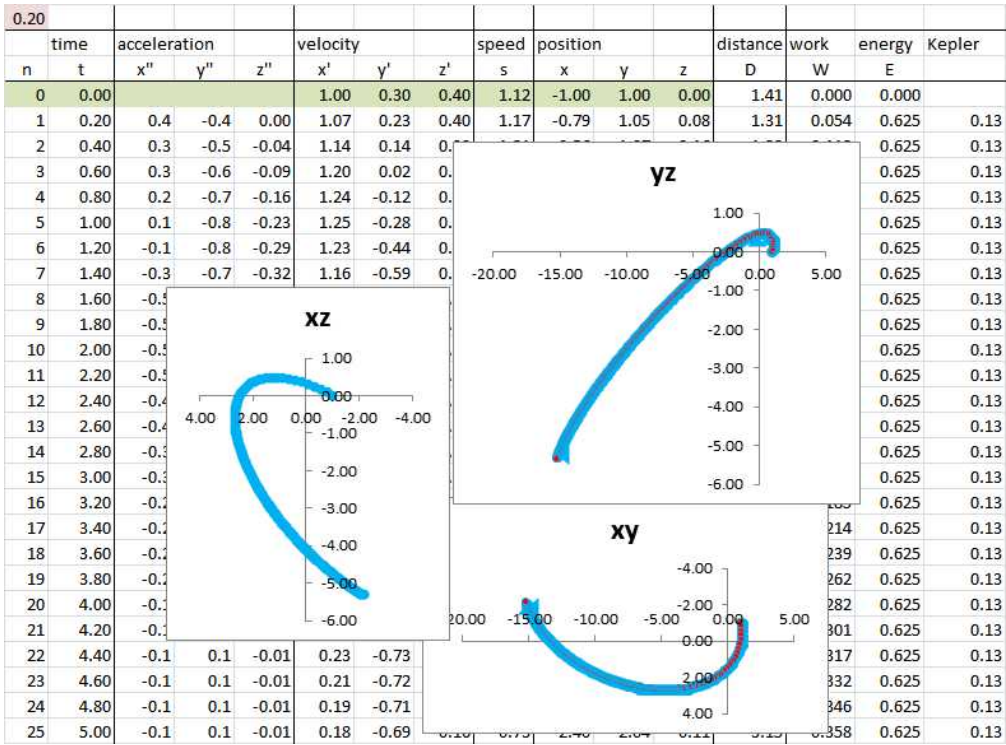
Finally, $A_{n+1} = A_n$ if and only if X_n and $f(X_n)$ are parallel.

Exercise 5.6.4

Derive Kepler’s Second Law from this theorem. Hint: use what you know about the accuracy of Euler’s method.

We thus conclude that the motion along the ellipse (or the parabola, or the hyperbola seen below) does *not* match its standard parametrization, $x = a \sin \omega t$, $y = b \sin \omega t$.

The formulas fully apply to the 3-dimensional situation. Plotting the orbit along the three coordinate planes produces the following:



Exercise 5.6.5

Implement a simulation of planetary motion in the 3-dimensional space. Demonstrate that the motion is planar.

We state the following without proof:

Theorem 5.6.6: Planetary Motion in Polar Coordinates

Any trajectory of the Newton's ODE of planetary motion is given by the following in polar coordinates:

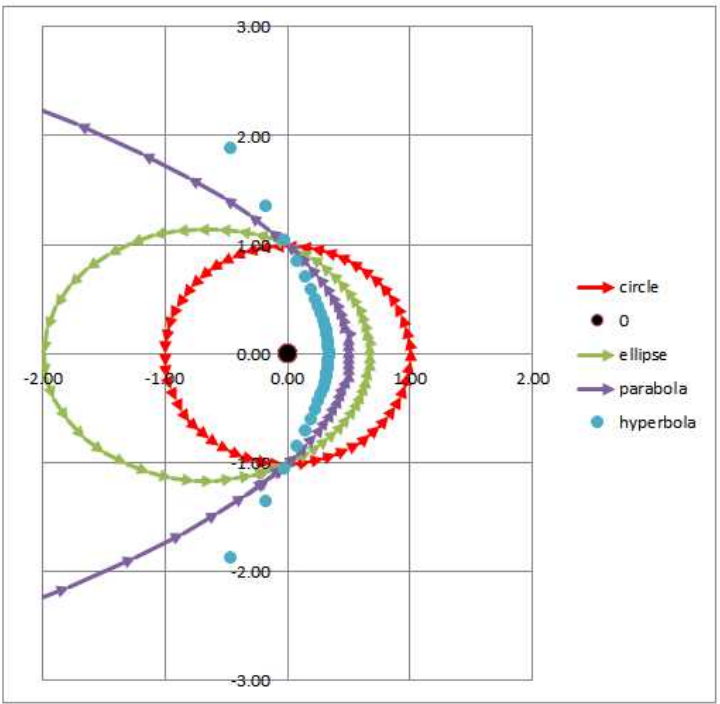
$$r = \frac{p}{1 + e \cdot \cos \theta},$$

which is

- a circle for $e = 1$,
- an ellipse for $0 < e < 1$,
- a parabola for $e = 1$,
- a hyperbola for $e > 1$,

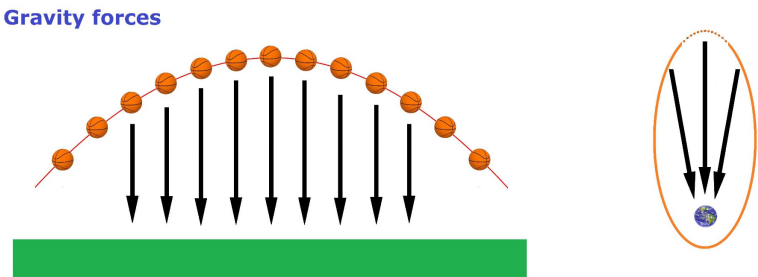
with the focus at the origin.

The graphs are plotted for $e = 0, .5, 1, 2$:



The proof of this theorem lies beyond the scope of this book.

Finally, let’s go back to the question posed in the beginning of the section: what is the correct trajectory?



Let’s review what we concluded about these two settings in [Chapter 4HD-2](#).

- When the Earth is seen as “large” in comparison to the size of the trajectory, the gravity forces are assumed to be parallel in all locations. Then the trajectory is a parabola as the graph of a function with the x -axis aligned with the surface of the Earth and y -axis with the force.
- When the Earth is seen as either “small” or at least perfectly spherical, the gravity forces are assumed to go radially toward its center. Then the trajectory is an ellipse (or a hyperbola or a parabola) with its focus located at the center.

When the size and, therefore, the shape of the Earth matter, things get complicated...

5.7. The two- and three-body problems

We have ignored the effect of the Earth’s gravity on the Sun. The reason is that the mass p of the Sun is significantly larger than the mass q of the Earth:

$$p \gg q.$$

This is the assumption that allows us to place the Sun at the origin as a stationary object. What if we have the *two planets of comparable sizes*?

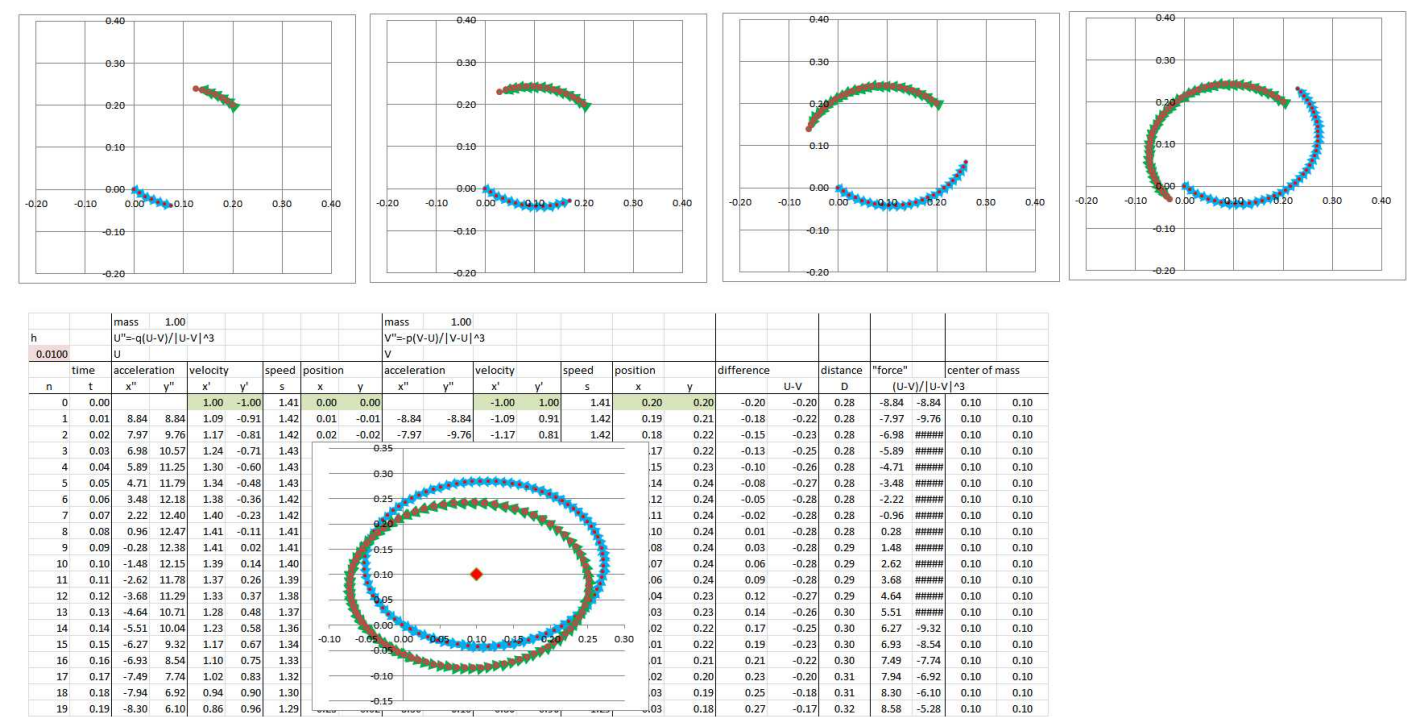
Suppose now we have two planets with masses p, q with located at U, V respectively. Either of the two objects is affected by the gravity of the other. Then the two interactions appear in the two dependent ODEs:

$$U'' = -Gq \frac{U - V}{||U - V||^3}$$
$$V'' = -Gp \frac{V - U}{||V - U||^3}$$

This is called the *two-body problem*.

Let’s apply Euler’s method to see what can happen.

We start with two identical planets. They seem to circle each other:

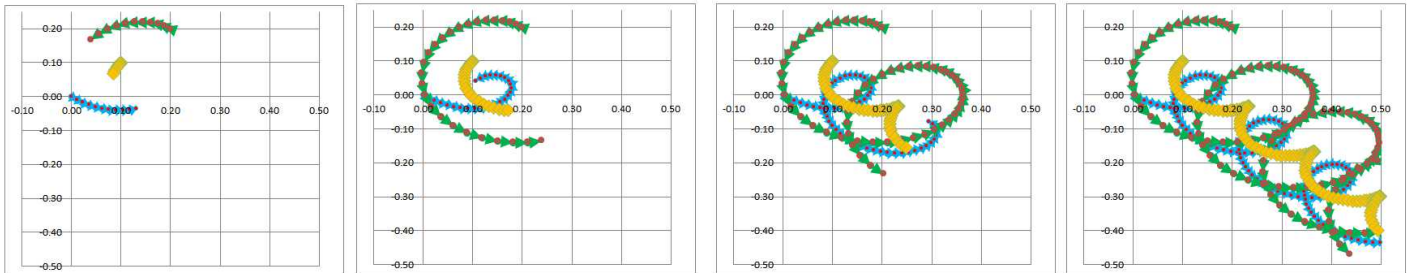


...until we realize that either one circles a certain point, the same point for both. This point lies half-way between then,

$$C = \frac{1}{2}(U + V) .$$

This fact is confirmed in the last column of the spreadsheet.

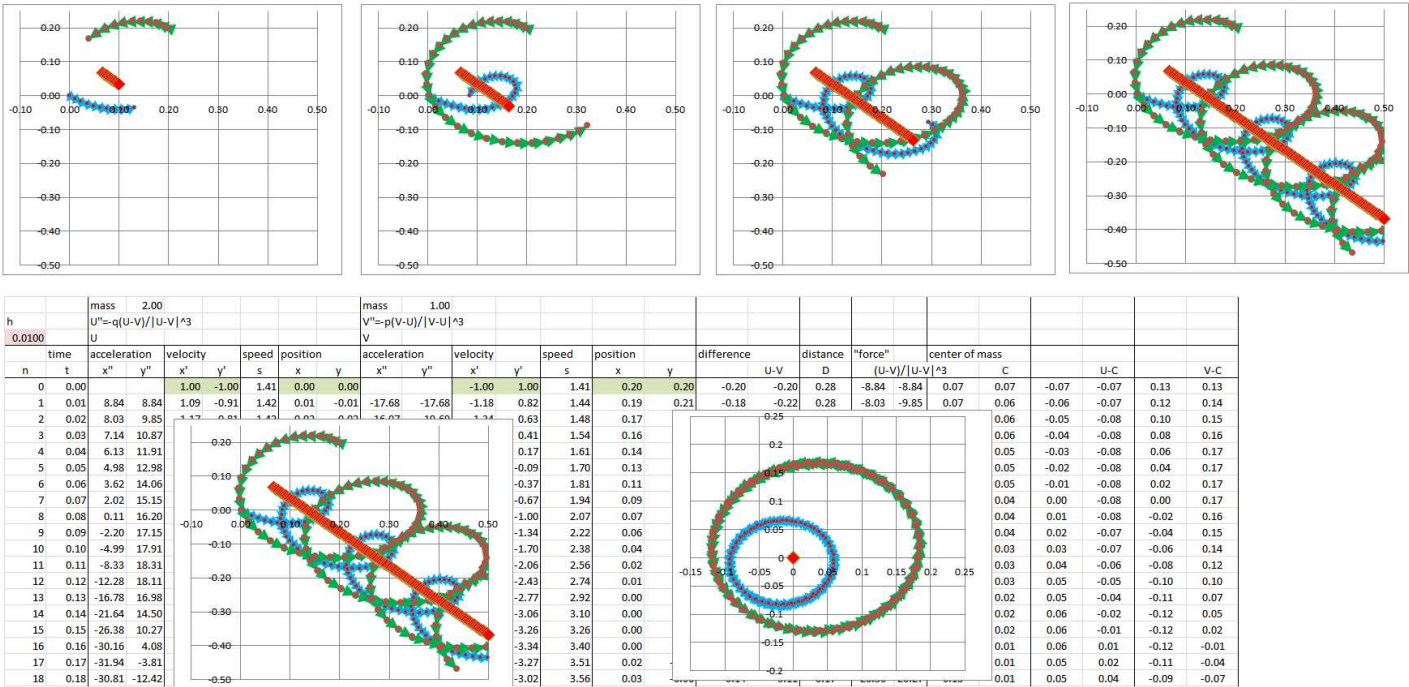
Let’s double the mass of the first planet. With everything else remaining the same, the two planets seem to dance away while still circling each other. This time, the mid-point is also circling.



In fact, it doesn’t seem to reveal anything about what is going on. What if we look at the *center of mass* of the two,

$$M = \frac{p}{p + q}U + \frac{q}{p + q}V ,$$

instead? It seems to be moving along a straight line!



Furthermore, let’s plot the two trajectories with respect to the center of mass, i.e.,

$$P = U - M \text{ and } Q = V - M.$$

They seem to trace ellipses.

Let’s state and prove these conjectures.

Theorem 5.7.1: Center of Mass Two-body System

The center of mass M of the two-body system satisfies the vector ODE:

$$M'' = 0,$$

and, therefore, moves along a straight line at a constant speed.

Proof.

We simply substitute:

$$\begin{aligned} M'' &= \left(\frac{p}{p+q}U + \frac{q}{p+q}V \right)'' \\ &= \frac{p}{p+q}U'' + \frac{q}{p+q}V'' \\ &= \frac{p}{p+q} \left(-Gq \frac{U-V}{||U-V||^3} \right) + \frac{q}{p+q} \left(-Gp \frac{V-U}{||V-U||^3} \right) \\ &= -G \frac{pq}{p+q} \left(\frac{U-V}{||U-V||^3} + \frac{V-U}{||U-V||^3} \right) \\ &= 0. \end{aligned}$$

Theorem 5.7.2: Locations in Two-body System

If U and V are the locations of the two planets in the two-body system and M is its center of mass M , then $P = U - M$ and $Q = V - M$ satisfy the Newton’s ODE of planetary motion and, therefore, trace ellipses, parabolas, or hyperbolas.

Proof.

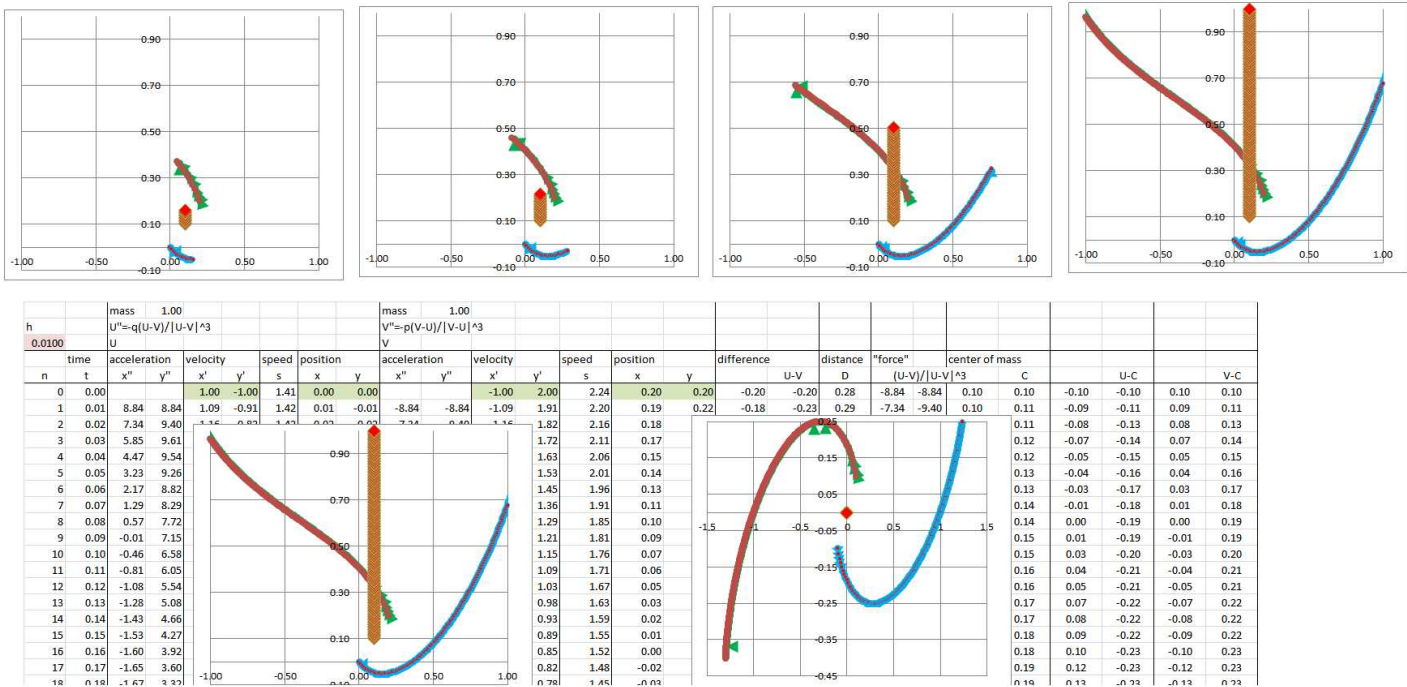
We first observe that M located on the line between U and V in proportion to their masses. Therefore, $U - V$ is proportional to $U - M$:

$$M = \frac{p}{p+q}U + \frac{q}{p+q}V \implies U - V = \frac{p+q}{q}(U - M).$$

Now, we simply substitute:

$$\begin{aligned}
 P'' &= (U - M)'' \\
 &= U'' - M'' \\
 &= -Gq \frac{U - V}{||U - V||^3} - 0 \\
 &= -Gq \frac{\frac{p+q}{q}(U - M)}{||\frac{p+q}{q}(U - M)||^3} \\
 &= -Gq \frac{U - M}{\left(\frac{p+q}{q}\right)^2 ||(U - M)||^3} \\
 &= -Gq \left(\frac{q}{p+q}\right)^2 \frac{U - M}{||(U - M)||^3} \\
 &= -G \frac{q^3}{(p+q)^2} \frac{P}{||P||^3}.
 \end{aligned}$$

In the last example, we go back to the case of two identical planets and simply increase the y -component of the velocity of the second planet from 1 to 2. With a certain amount of circling, they start to move away from each other.



Eventually, they are flying in the opposite directions! Just as in the last case, plotting the trajectories with respect to the center of mass help to reveal the pattern.

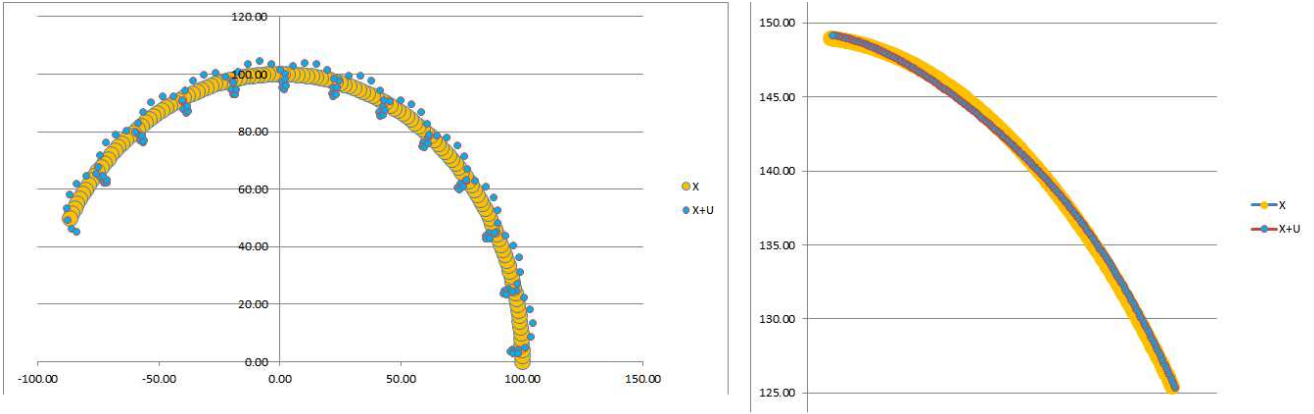
Thus, depending on the initial conditions, these three seem to be the most interesting outcomes. In the first, two identical planets *dance around* each other (in reality, around the combined center of mass). In the second, they *dance away together* (still around the center of mass, which is moving). Finally, the two *run away from each other* because beyond some distance they feel almost no pull...

Exercise 5.7.3

Why don't any of the planets in the solar system behave this way? What other possibilities can you think of?

What about *three planets*?

First let's review a related problem discussed in [Chapter 4HD-2](#). The fact that the Earth orbits the Sun and the Moon orbits around the Earth may be illustrated with the picture on left while the actual data about the dimensions of the orbits and the period of the Moon produce the picture on right:



Suppose now that the three planets have masses p, q, m . If we assume that

$$p \gg q \gg m,$$

such as in the case of Sun-Earth-Moon, the effect of the gravity of the second on the first and the third on the second is negligible, just as in the last section. Suppose

- $U = 0$ is the location of the Sun fixed at the origin,
- V is the location of the Earth with respect to the Sun, and
- X is the location of the Moon with respect to the Earth.

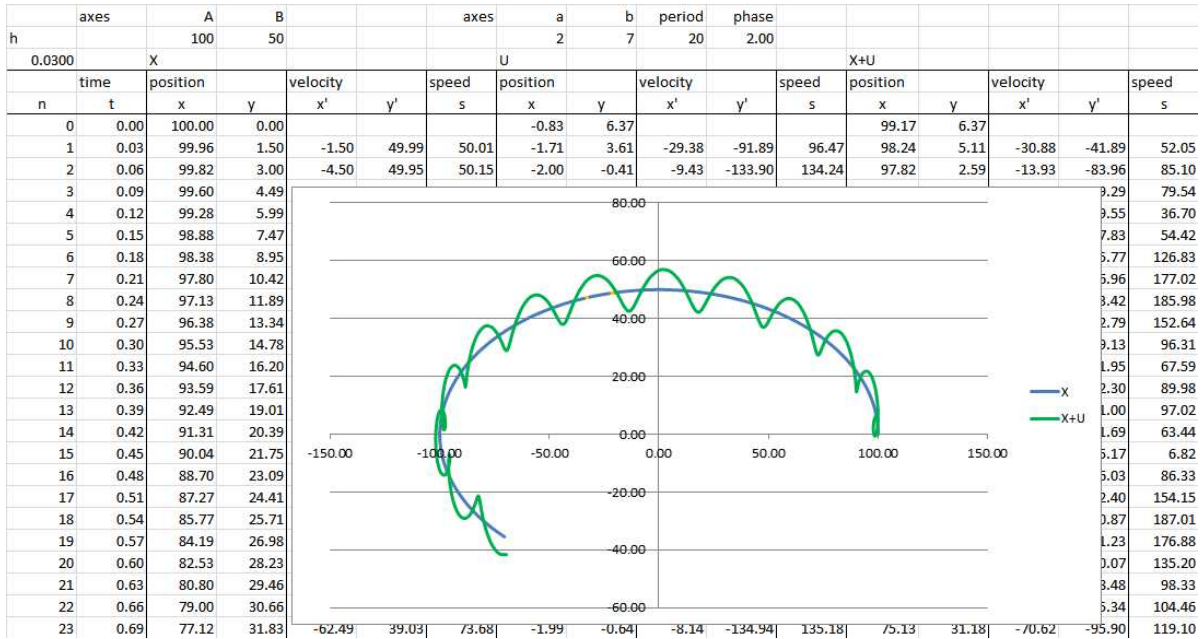
Then the two interactions are independent of each other and we have two independent Newton's ODEs of planetary motion:

$$\begin{aligned} V'' &= -Gp \frac{V}{||V||^3} \\ X'' &= -Gq \frac{X}{||X||^3} \end{aligned}$$

Both solutions follow their respective ellipses. We then add them as vectors,

$$W = V + X,$$

to produce the path of the third planet around the first. Generically, it looks like this:

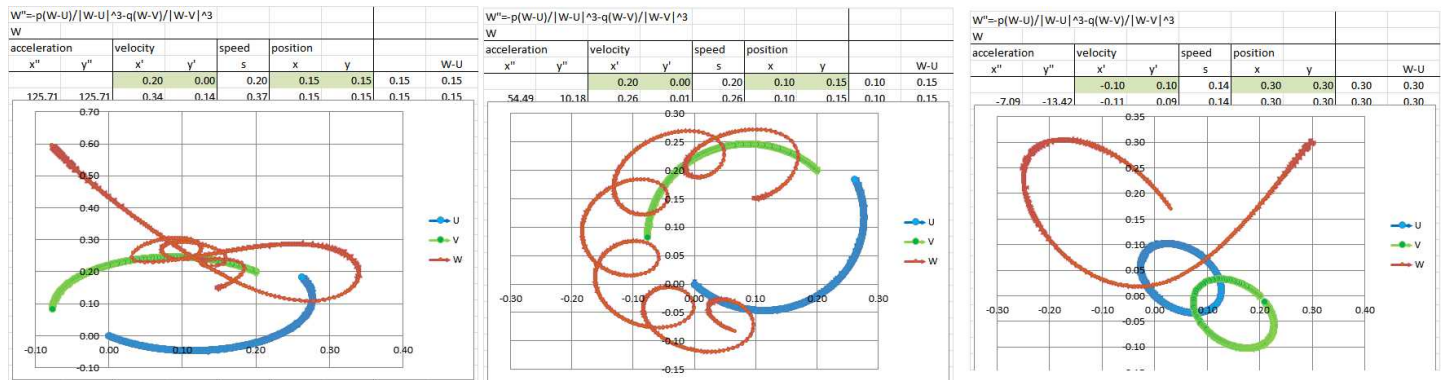


We next make one generalization: the mass of the second planet is *not negligibly small* relative to that of the first anymore. We, then, can't fix the first at the origin anymore. We have a two-body system just as above:

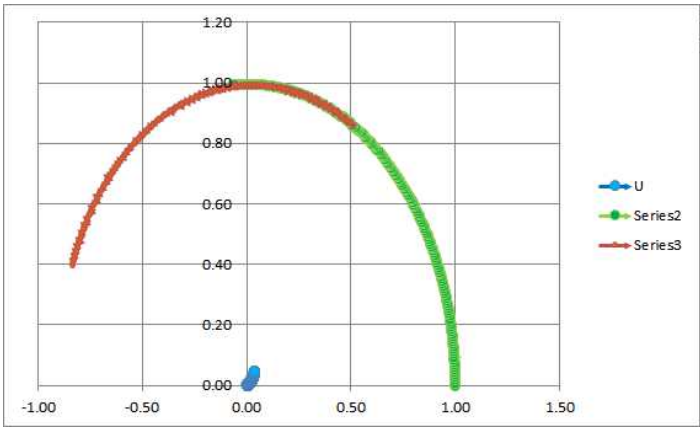
$$U'' = -Gq \frac{U - V}{\|U - V\|^3}$$
$$V'' = -Gp \frac{V - U}{\|V - U\|^3}$$
$$W'' = -Gp \frac{W - U}{\|W - U\|^3} - q \frac{W - V}{\|W - V\|^3}$$

In the meantime, the third planet is affected by the gravity of the first two. It is called the *restricted three body problem*

While the two first planets do the same, the variety of behaviors of the third is enormous even in dimension 2:



There are many periodic trajectories but we will point out only one. Below we confirm that it is possible to place a satellite on the Moon's orbit, 60 degrees ahead, that will continue to revolve in this fashion indefinitely.



Now, in general. Suppose now we have three planets with masses m_1, m_2, m_3 located at V_1, V_2, V_3 respectively. Each of the three objects is affected by the gravity of either of the other two. Then these three interaction appear in the three dependent vector ODEs:

$$V_1'' = -Gm_2 \frac{V_1 - V_2}{||V_1 - V_2||^3} - Gm_3 \frac{V_1 - V_3}{||V_1 - V_3||^3}$$
$$V_2'' = -Gm_1 \frac{V_2 - V_1}{||V_2 - V_1||^3} - Gm_3 \frac{V_2 - V_3}{||V_2 - V_3||^3}$$
$$V_3'' = -Gm_1 \frac{V_3 - V_1}{||V_3 - V_1||^3} - Gm_2 \frac{V_3 - V_2}{||V_3 - V_2||^3}$$

The general three body problem has no analytic solution.

5.8. A cannon is fired...

What we know about history of ballistics is that the parabolic trajectories were known since Galileo and, furthermore, Newton worked out the physics behind and the mathematics (calculus) behind this idea. Nonetheless, as recently as mid-19th century they aimed cannons based entirely on the information that comes from experimentation, without mathematics.

Here is an excerpt from *The Emperor Napoleon's New System of Field Artillery* (1854):

XXVII. The elevations to be given to the different pieces for practice beyond point blank range, are given in the following table.

Nature of Ordnance.		Ranges in metres. Elevations in millimetres.									
		met. 300	met. 400	met. 500	met. 600	met. 700	met. 800	met. 900	met. 1000	met. 1100	met. 1200
		m.m.	m.m.	m.m.	m.m.	m.m.	m.m.	m.m.	m.m.	m.m.	m.m.
Howitzers.	12-pound guns	4	14	23	33	48	61	75
	8-pound guns	10	19	29	41	56	74	92*
	16 ^c large charge	7	20	33	49	65	84	108	128*
	„ small charge	7	21	38	56	76	97*	117*	138*
	15 ^c large charge	...	4	16	27	41	56	74	92*	111*	133*
	„ small charge	7	21	34	48	62	78	96*	114*

Note.—The elevations marked with an asterisk are obtained by interpolations, the others are the result of experiment.

Below is the “range table” from the *The Confederate Ordnance Manual* during the American Civil War (1860s). Clearly, the numbers come from shooting and then measuring the distance:

Description of pieces	Charge in pounds	Projectile Type	Elevation in degrees & minutes	Range in yards	Remarks
12 - PDR. FIELD HOWITZER	1	Shell	0	195	
		"	1	539	
		"	2	640	
		"	3	847	
		"	4	975	
		"	5	1072	
	0.75	Sph-case	2 15	485	Time, 3 Seconds " 4 "
		"	3 15	715	
		"	3 45	1,050	

With what we know about the dynamics of projectiles, we can try to reproduce these results. We start with the same initial value problem:

$$\begin{cases} x'' = 0 \\ y'' = -32 \end{cases}, \quad \begin{cases} x'(0) = s_0 \cos \alpha \\ y'(0) = s_0 \sin \alpha \end{cases}, \quad \begin{cases} x(0) = 0 \\ y(0) = y_0 \end{cases},$$

where s_0 is the initial speed and α is the angle of the barrel.

Example 5.8.1: data

The initial value problem has already been solved:

$$\begin{cases} x = s_0 \cos \alpha \cdot t, \\ y = y_0 + s_0 \sin \alpha \cdot t - 16t^2. \end{cases}$$

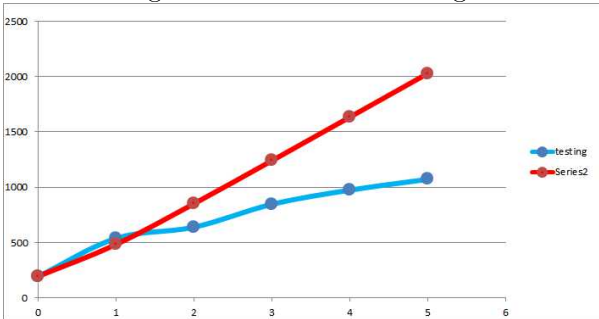
We need to find such an x that $y = 0$. We find the time to reach the ground first (choosing the positive value):

$$t = \frac{-s_0 \sin \alpha \pm \sqrt{(s_0 \sin \alpha)^2 - 4(-16)y_0}}{2(-16)} = \frac{1}{32} \left(s_0 \sin \alpha + \sqrt{(s_0 \sin \alpha)^2 + 64y_0} \right).$$

We consider the 12-pound field howitzer. Suppose the muzzle velocity is known, $s_0 = 1054$ feet per second. We also estimate the initial elevation (the height of the cannon) at $y_0 = 5$ feet. Then the distances should be:

elevation in degrees	time in seconds, t	distance in feet, x	distance in yards	test distance
0	0.56	589	196	195
1	1.38	1451	484	539
2	2.43	2557	852	640
3	3.54	3722	1241	847
4	4.66	4902	1634	975
5	5.80	6085	2028	1072

We see a mismatch with the data that grows fast with the angle.



Exercise 5.8.2

Explain why the red curve looks straight.

The reason why the Newtonian mechanics overestimates the distance is known; it is the *air-resistance*. In fact, Newton himself worked out all necessary details of ballistics including taking into account the air resistance. He assumed however that the resistance force is proportional to the speed. Half a century later Euler thought, for ballistics, it is better to take the resistance force is proportional to the *square* of the speed. That was 100 years before those tests!

The *drag equation* gives the force F_D of drag experienced by an object due to its movement through the air:

$$||F_D|| = \tfrac{1}{2}\rho C A \cdot s^2 \, ,$$

where

- ρ is the density of the air,
- A is the cross sectional area of the projectile,
- C is the drag coefficient – a dimensionless coefficient that depends on the projectile’s geometry, and
- s is the speed of the projectile.

The direction of this force F_D and, therefore, that of the acceleration is opposite to the direction of the velocity X' . Therefore, the acceleration generated by this force is

$$-\tfrac{1}{2}\rho C A \tfrac{1}{M} s X' \, ,$$

where M is the mass of the projectile.

Then our IVP becomes:

$$\begin{cases} x'' &= & -\tfrac{1}{2}\rho C A \tfrac{1}{M} s x' \\ y'' &= & -32 -\tfrac{1}{2}\rho C A \tfrac{1}{M} s y' \end{cases} \, , \quad \begin{cases} x'(0) &= & s_0 \cos \alpha \\ y'(0) &= & s_0 \sin \alpha \end{cases} \, , \quad \begin{cases} x(0) &= & 0 \\ y(0) &= & y_0 \end{cases} \, .$$

With this updated dynamics of projectiles, let’s try again to reproduce the test results.

Example 5.8.3: data

We need some information for the drag equation:

- the weight $M = 8.9$ pounds
- the area $A = \pi r^2$ of the cross section of the cannonball, where the radius is $r = 4.62/2 = 2.31$ inches
- the density of the air $\rho = 0.074887$ pounds per cubic foot
- the drag coefficient of a sphere $C = 0.47$

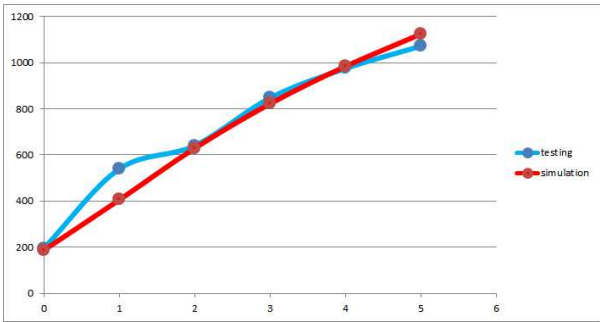
We use Euler’s method:

1/2rCA=	0.00205	bore:	4.62 in	range:	3373.18 ft=	1124 yards													
muzzle vel.:	1054 ft/sec	weight:	8.9 lbs			test:	1072												
angle:	5.0 deg					error:	0.05 %												
		y''=-32-y's																	
0.02																			
	time	acceleration	velocity		speed	position													
	n	t	x"	y"	x'	y'	s	x	y										
	0	0.00			1049.99	91.86	1054.00	0.00	5.00										
	1	0.02	-254.8	-54.3	1046.17	91.05	1050.12	15.69	6.37										
	2	0.03	-252.9	-54.0	1042.37	90.24	1046.27	31.33	7.72										
	3	0.05	-251.1	-53.7	1038.61	89.43	1042.45	46.91	9.06										
	4	0.06	-249.2	-53.5	1034.87	88.63	1038.66	62.43	10.39										
	5	0.08	-247.4	-53.2	1031.16	87.83	1034.89	77.90	11.71										
	6	0.09	-245.6	-52.9	1027.47	87.04	1031.15	93.31	13.01										
	7	0.11	-243.9	-52.7	1023.82	86.25	1027.44	108.67	14.31										
	8	0.12	-242.1	-52.4	1020.18	85.46	1023.76	123.97	15.59										
	9	0.14	-240.4	-52.1	1016.58	84.68	1020.10	139.22	16.86										
	10	0.15	-238.7	-51.9	1013.00	83.90	1016.47	154.41	18.12										
	11	0.17	-237.0	-51.6	1009.44	83.13	1012.86	169.56	19.36										
	12	0.18	-235.4	-51.4	1005.91	82.36	1009.28	184.64	20.60										

We repeat it for each elevation, six times:

elevation in degrees	distance in yards	test distance
0	187	195
1	405	539
2	629	640
3	822	847
4	984	975
5	1124	1072

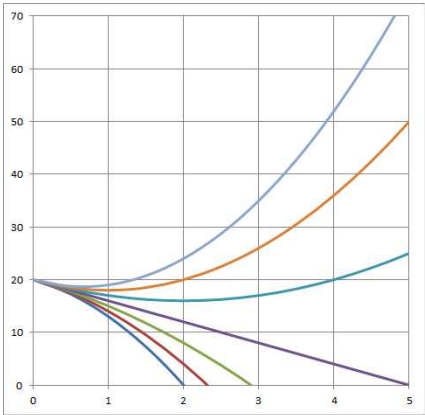
Much better!



5.9. Boundary value problems

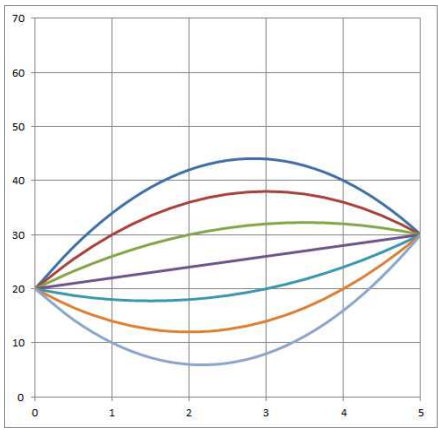
Example 5.9.1: sample BVP

Setting differential equations aside for a moment, let's consider this simple problem:
► From all quadratic polynomials $f(x) = ax^2+bx+c$, find the ones with $f(0) = 20$, $f'(0) = -3$.
Here, the value of the function and its derivative are provided for a particular value of the variable, $x = 0$. These are called *initial conditions*. The word “initial” refers to the starting point, 0, of the ray $[0, \infty]$.



initial conditions
 $f(0)=20$
 $f'(0)=-3$

Let's change the problem:
► From all quadratic polynomials $f(x) = ax^2+bx+c$, find ones with $f(0) = 20$, $f(5) = 30$.
In this case, the value of the function is provided for two values of x and no values of the derivative are provided. These are called *boundary conditions*. The word “boundary” refers to the end-points, 0 and 5, also known as the boundary points, of the interval $[0, 5]$.



boundary conditions

$f(0)=20$
 $f(5)=30$

Exercise 5.9.2

Find all these polynomials.

Example 5.9.3: unitary quadratic polynomials

Let’s limit ourselves to quadratic polynomials with $a = 1$:

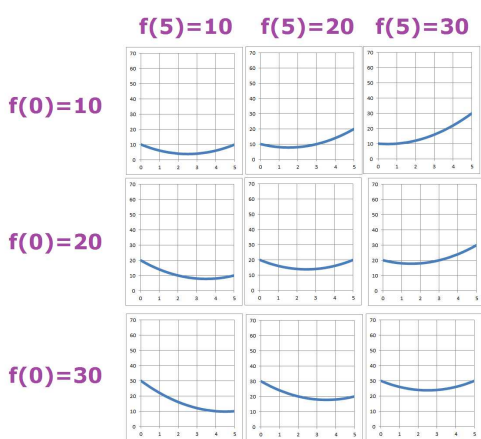
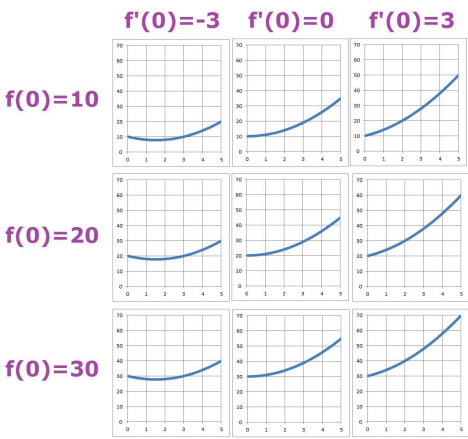
- From all quadratic polynomials $f(x) = x^2 + bx + c$, find the one with $f(0) = 20$, $f'(0) = -3$.
- From all quadratic polynomials $f(x) = x^2 + bx + c$, find the one with $f(0) = 20$, $f(5) = 30$.

Then we have a certain *uniqueness*, i.e., there is only one answer to the question; the solution in the former case is, of course, $c = 20$, $b = -3$. For the latter case we need more algebra:

$c = 20 \implies 5^2 + b \cdot 5 + 20 = 30 \implies b = (30 - 5^2 - 20)/5 = 3.$

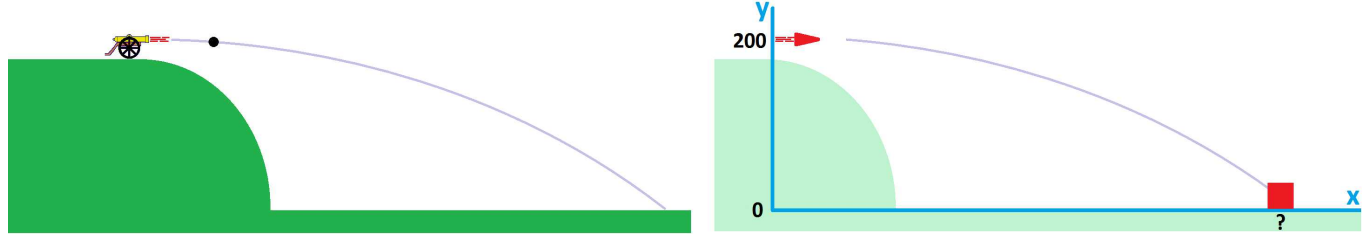
initial conditions

boundary conditions



Example 5.9.4: cannon

Let’s review a familiar problem from Volumes 1 and 2. From a 200 feet elevation, a cannon is fired horizontally at 200 feet per second. How far will the cannonball go?



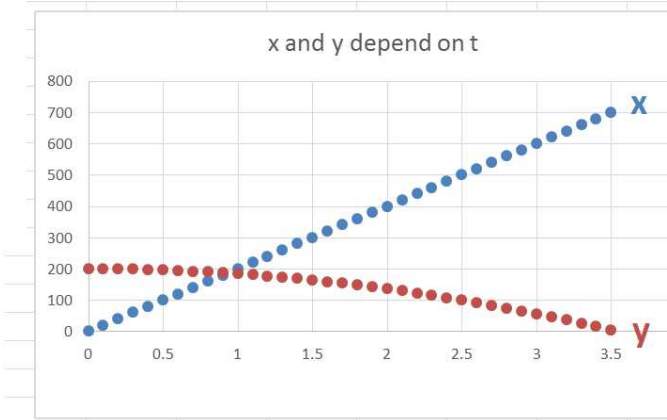
We know this as an initial value problem. Indeed, the ODE has been solved:

$$\begin{cases} x = 200t, \\ y = 200 - 16t^2. \end{cases}$$

Now we just need to find the solution that satisfies the *initial conditions*:

- the initial location: $X(0) = (x(0), y(0)) = (0, 200)$;
- the initial velocity: $X'(0) = (x'(0), y'(0)) = (200, 0)$.

time	depth x	height y
0	0	200
0.1	20	199.84
0.2	40	199.36
0.3	60	198.56
0.4	80	197.44
0.5	100	196
0.6	120	194.24
.....		
3.2	640	36.16
3.3	660	25.76
3.4	680	15.04
3.5	700	4
3.6	720	-7.36



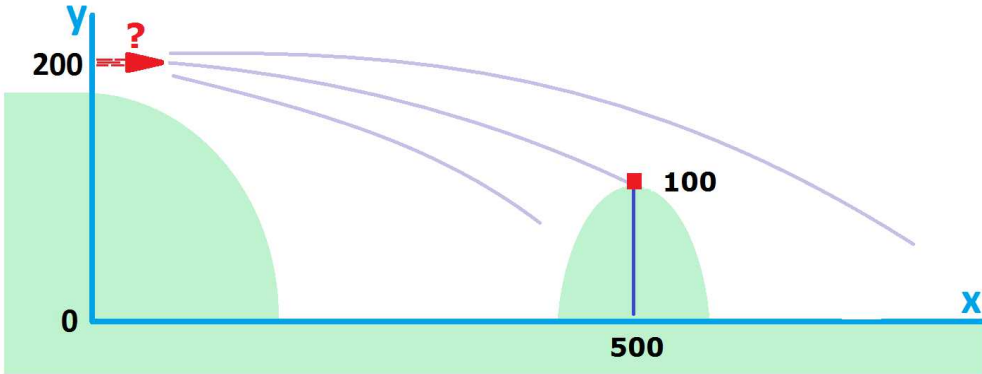
We scroll down the spreadsheet to find the row with y close to 0: around $t_1 = 3.55$ seconds, with the value of x at the time close to 710 feet. Algebraically, we solve an equation and then substitute:

$$y(t_1) = 200 - 16t_1^2 = 0 \implies t_1 = \sqrt{\frac{200}{16}} = \frac{5\sqrt{2}}{2} \implies x_1 = x(t_1) = 200t_1 = 200\frac{5\sqrt{2}}{2} \approx 707.$$

Now what could be the meaning of boundary conditions in this setting? We already have one, the *initial location*. The second may be the *final location* when, for example, we are trying to hit a target – at a particular moment of time. Suppose we want the cannon ball to be at $(200, 500)$ after 2 seconds. In that case the muzzle velocity of the cannon is unspecified and it is what we are supposed to find. We thus need to find a solution (there could be one, many, or none) that satisfies the *boundary conditions*:

- the first boundary location: $X(0) = (x(0), y(0)) = (0, 200)$;
- the second boundary location: $X(2) = (x(2), y(2)) = (500, 100)$.

This is called a *boundary value problem*. Let’s solve it.



Where do the solutions come from? The ODE has been solved and the first boundary condition gives us the initial location:

$$\begin{cases} x = x_0 + ut \\ y = y_0 + vt - 16t^2 \end{cases} \implies \begin{cases} x = 0 + ut \\ y = 200 + vt - 16t^2 \end{cases}.$$

We now find u, v from the second boundary condition.

One dimension at a time: for x ,

$$x(2) = v \cdot 2 = 500 \implies u = 250;$$

and for y ,

$$y(2) = 200 + v \cdot 2 - 16 \cdot 2^2 = 100 \implies 2v = 100 - (200 - 16 \cdot 2^2) \implies v = (-100 + 16 \cdot 4)/2 = -18.$$

Note that a more practical situation is when the muzzle velocity, mathematically the *speed*, remains the same and it is the *angle* of the barrel that we need to find.

Example 5.9.5: spring

Consider the familiar ODE for the oscillation of an object on a spring:

$$y''(x) + y(x) = 0 .$$

But this time we aren't trying to predict what happens to the object with known position and velocity. We ask ourselves if the system can bring the object from a particular position to another in a specific amount of time, for example:

$$y(0) = 0, \; y(\pi/2) = 2 .$$

The general solution to our ODE is

$$y(x) = A \sin x + B \cos x .$$

Now, from the first boundary condition we obtain:

$$0 = A \cdot 0 + B \cdot 1 \implies B = 0 .$$

From the boundary condition we obtain:

$$2 = A \cdot 1 \implies A = 2 .$$

Thus imposing these boundary conditions produces a unique solution:

$$y(x) = 2 \sin x .$$

When the functions have two or more variables, regions are more complicated and so are their boundaries. Such differential equations are beyond the scope of this book.

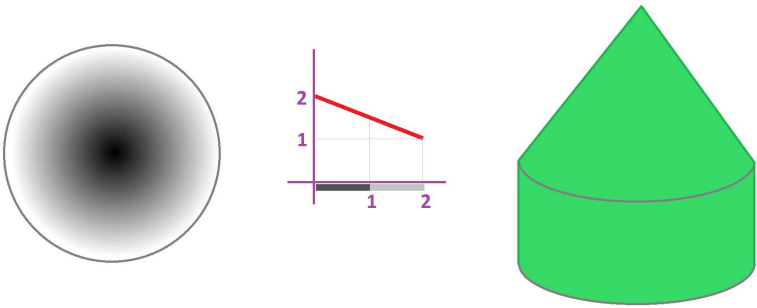
Chapter 6: Partial differential equations

Contents

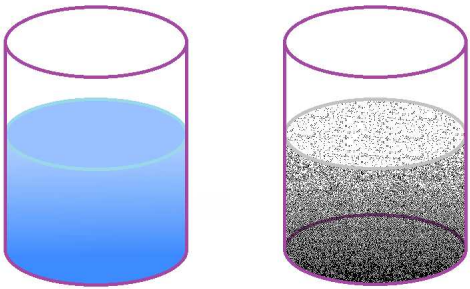
6.1 Heat transfer between adjacent objects	343
6.2 Heat transfer depends on permeability	353
6.3 Heat transfer is caused by temperature differences	360
6.4 Heat transfer depends on the geometry	365
6.5 The heat PDE	370
6.6 Cells and forms in higher dimensions	375
6.7 Heat transfer in dimension 2: a plate	383
6.8 The heat PDE for dimension 2	391
6.9 Wave propagation in dimension 1: springs and strings	393
6.10 The wave PDE	400
6.11 Wave propagation in dimension 2: a membrane	405

6.1. Heat transfer between adjacent objects

Motion happens in time and space. However, you can't be at two places at the same time! In this chapter, we will investigate processes that happen – separately but not independently – at every location of a region. A cup may be gradually cooling from the outside:



The temperature varies with time at every location but also from location to location at every moment of time. Or it may be warming up on a stove:



The patterns of the distribution of temperature in the cup differ in these two scenarios.

We start with a review and then proceed to the 1-dimensional case.

Let’s recall Newton’s Law of Cooling (Chapter 1):

► “The rate of cooling is proportional to the difference of the current temperature and the room temperature”.

An example is cooling of a cup of coffee or warming up a can of soda.

We introduce variables:

- 1. t is the time.
- 2. u is the temperature.

We then rewrite the description in terms of the difference quotient:

$$\frac{\Delta u}{\Delta t} = k(r - u) ,$$

where r is the room temperature, for some constant $k > 0$. We can see that when the temperature is higher than the room temperature, it decreases and when the temperature is lower than the room temperature, it increases.

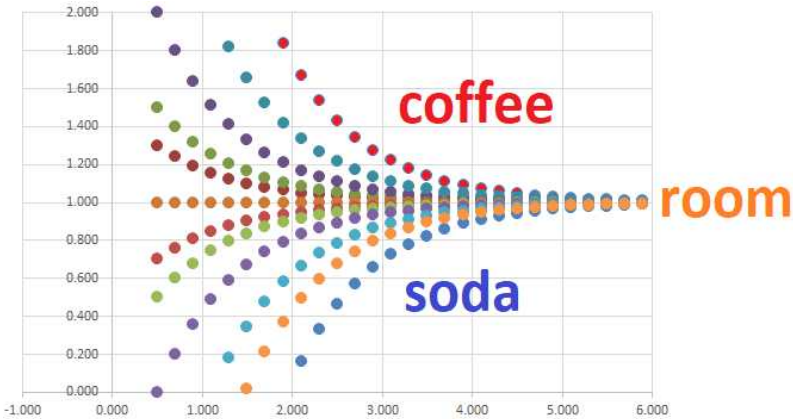
The equation gives us a solution via this recursive formula (difference equation):

$$u(t_{n+1}) = u(t_n) + K(r - u(t_n)) \cdot \Delta t ,$$

combined with

$$t_{n+1} = t_n + \Delta t .$$

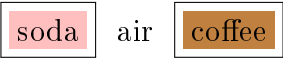
These formulas are used for simulations. We can see below several possible solutions. In particular, a cup of coffee is cooling down and a can of soda is warming up:



Both temperatures are approaching that of the room as they exchange the heat with the surrounding air according to the theorem in Chapter 1. The room temperature remains constant because the amount of heat is too large in comparison.

We proceed as follows.

We first take this model as that of *two objects* in the room at the same time:



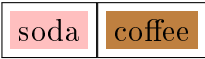
We have two functions:

u is the temperature of the soda and v is the temperature of the coffee.

Then, we have two recursive formulas independent of each other:

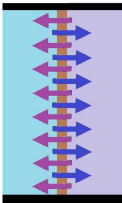
$$u(t_{n+1}) = u(t_n) + K(r - u(t_n)) \cdot \Delta t \quad \text{and} \quad v(t_{n+1}) = v(t_n) + K(r - v(t_n)) \cdot \Delta t.$$

Next step: What if the two objects are *adjacent* to each other? We assume that the two touch each other while insulated elsewhere (no air):



Then the either of the two objects exchanges the heat with the other and nothing else. The law still applies:

- “The rate of heat exchange is proportional to the difference of the current temperature and the temperature of the adjacent object”.



At each step, therefore, the dynamics remains the same as before but the reference is to the temperature of the other object rather than the room temperature.

We take the two formulas formulas above and just replace r (air) with the other object’s temperature:

- for soda u : r replaced with coffee v ,
- for coffee v : r replaced with soda u .

This is the result:

$$u(t_{n+1}) = u(t_n) + k(v(t_n) - u(t_n)) \cdot \Delta t \quad \text{and} \quad v(t_{n+1}) = v(t_n) + k(u(t_n) - v(t_n)) \cdot \Delta t.$$

We can see that the heat added to one is the heat taken from the other; that *Conservation of Energy*!

Example 6.1.1: soda-coffee

Let’s carry out this plan with a *spreadsheet*.

We choose the initial temperatures: soda 40 and coffee 100 degrees; they are in the first row. We chose $k = .1$ for a slow exchange. The two formulas are similar:

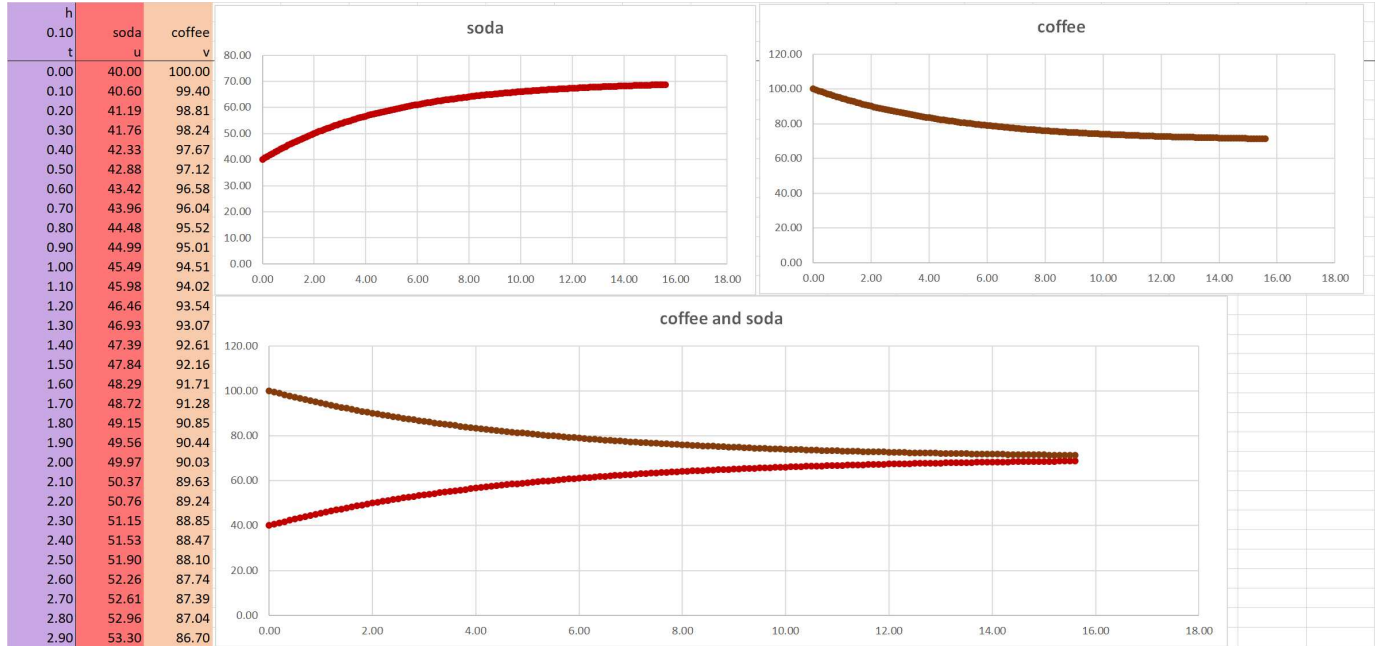
`=R[-1]C+0.1*(R[-1]C[1]-R[-1]C)*R2C` and `=R[-1]C+0.1*(R[-1]C[-1]-R[-1]C)*R2C`

Below is the precedence dependence of these formulas:

h	soda	coffee
t	u	v
0.00	40.00	100.00
0.10	40.60	99.40
0.20	41.19	98.81

h	soda	coffee
t	u	v
0.00	40.00	100.00
0.10	40.60	99.40
0.20	41.19	98.81

As expected, the dynamics is almost identical to the original:



The difference is that the temperatures converge to the middle instead to the room temperature:

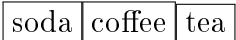
$$\frac{40 + 100}{2} = 70.$$

Exercise 6.1.2

Prove the limit.

One step further. What if there also *a cup of tea*?

We assume again that the three touch each other while insulated elsewhere:



The functions are respectively:

u *v* *w*

Then the soda exchanges heat with the coffee just as before and the tea also exchanges heat with the coffee just as before. Either has only one neighbor! We write for these two:

u(*t*_{*n*+1}) = *u*(*t*_{*n*}) + *k*(*v*(*t*_{*n*}) − *u*(*t*_{*n*})) · Δ*t*

w(*t*_{*n*+1}) = *w*(*t*_{*n*}) + *k*(*v*(*t*_{*n*}) − *w*(*t*_{*n*})) · Δ*t*.

The coffee is now in a new position. It has *two* neighbors! Since we expect the heat added to the coffee is the heat taken from the two others, we simply copy the terms from the above formulas. Therefore, we have:

v(*t*_{*n*+1}) = *v*(*t*_{*n*}) + *k*(*u*(*t*_{*n*}) − *v*(*t*_{*n*})) · Δ*t* + *k*(*w*(*t*_{*n*}) − *v*(*t*_{*n*})) · Δ*t*.

Example 6.1.3: soda-coffee-tea

We choose the initial temperatures: soda 40, coffee 100, and tea 85 degrees. The two formulas the soda and the tea are the same as before:

=R[-1]C+0.1*(R[-1]C[1]-R[-1]C)*R2C and =R[-1]C+0.1*(R[-1]C[-1]-R[-1]C)*R2C

They take from their right and left neighbors respectively.

The formulas the coffee is new:

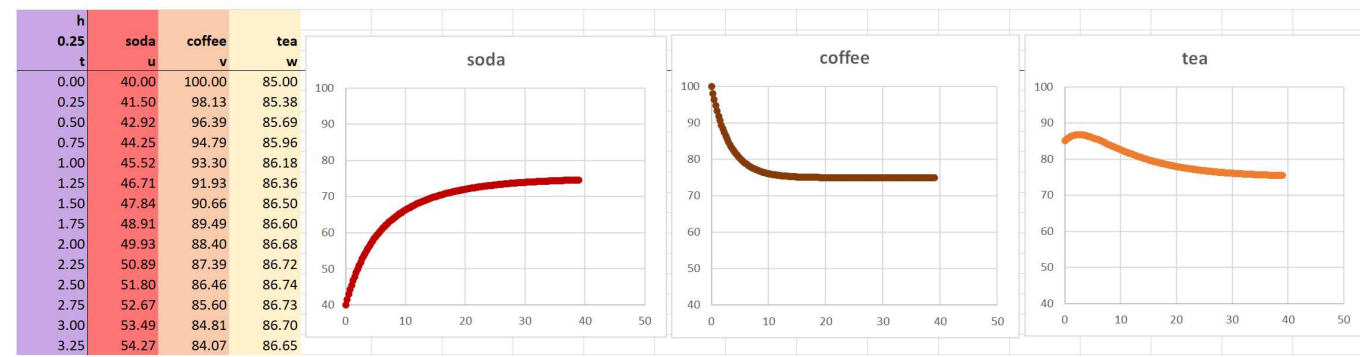
=R[-1]C+0.1*(R[-1]C-R[-1]C[-1])*R2C1+0.1*(R[-1]C-R[-1]C[1])*R2C1

It takes from their left and right neighbors at the same time.

These are the first rows:

h	soda		coffee		tea	
	t	u	v	w		
0.00	0.00	40.00	100.00	85.00		
0.25	0.25	41.50	98.13	85.38		

The dynamics is similar to the last:



The soda is warming up, the coffee is cooling down, while the tea first is warming up (because the adjacent coffee is hotter) and then cooling down (as the coffee is). They are all converging on the average temperature:

$$\frac{40 + 100 + 85}{3} = 75.$$

Next step: We imagine that there is a series of objects that exchange their heat with their neighbors.

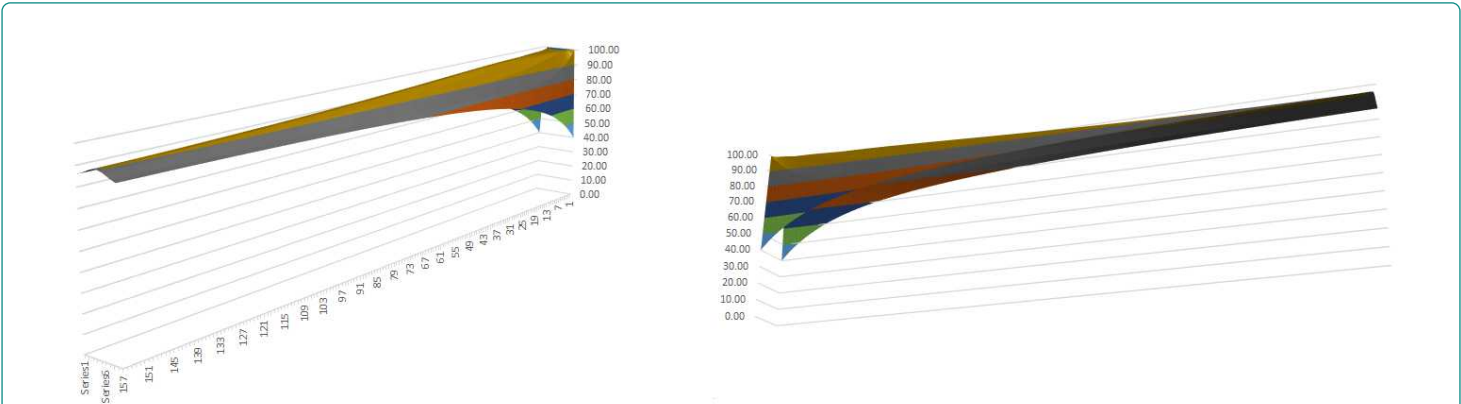
Example 6.1.4: ten cans

Suppose we have 10 objects (or containers) with cold ends and hot center. The last formula is repeated (eight times) in a spreadsheet with 10 columns. We color the cells to see the dynamics:

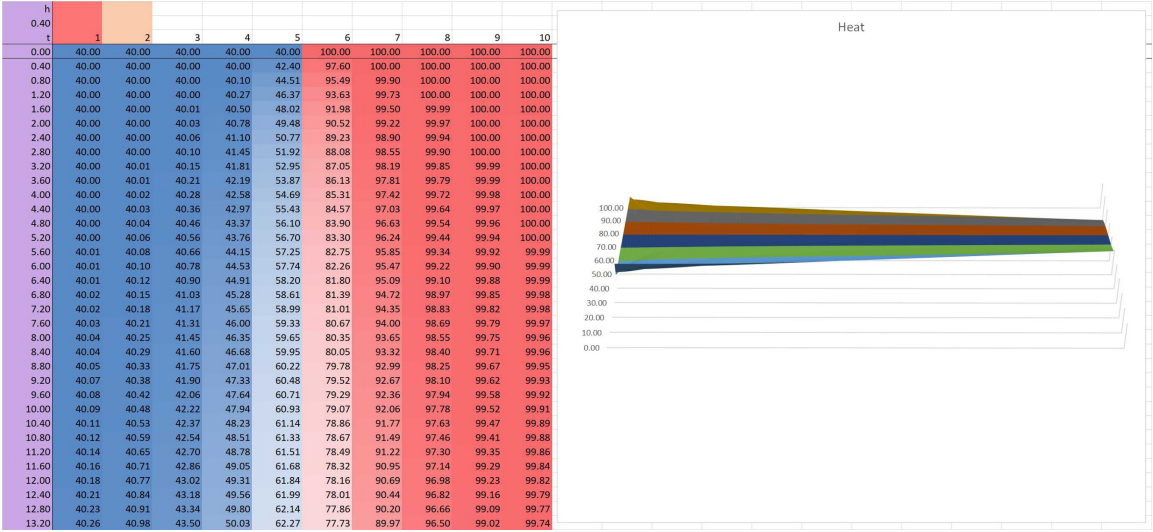
h	1	2	3	4	5	6	7	8	9	10
t										
0.00	40.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	40.00
0.40	42.40	97.60	100.00	100.00	100.00	100.00	100.00	100.00	97.60	42.40
0.80	44.61	95.49	99.90	100.00	100.00	100.00	100.00	99.90	95.49	44.61
1.20	46.64	93.63	99.73	100.00	100.00	100.00	100.00	99.73	93.63	46.64
1.60	48.52	91.99	99.50	99.99	100.00	100.00	99.99	99.50	91.99	48.52
2.00	50.26	90.56	99.22	99.97	100.00	100.00	99.97	99.22	90.56	50.26
2.40	51.87	89.29	98.90	99.94	100.00	100.00	99.94	98.90	89.29	51.87
2.80	53.37	88.18	98.56	99.90	100.00	100.00	99.90	98.56	88.18	53.37
3.20	54.76	87.20	98.20	99.85	99.99	99.99	99.85	98.20	87.20	54.76
3.60	56.06	86.34	97.82	99.79	99.99	99.99	99.79	97.82	86.34	56.06
4.00	57.27	85.59	97.44	99.72	99.98	99.98	99.72	97.44	85.59	57.27
4.40	58.40	84.93	97.06	99.64	99.97	99.97	99.64	97.06	84.93	58.40
4.80	59.47	84.36	96.68	99.55	99.95	99.95	99.55	96.68	84.36	59.47
5.20	60.46	83.85	96.30	99.45	99.94	99.94	99.45	96.30	83.85	60.46
5.60	61.40	83.42	95.93	99.34	99.92	99.92	99.34	95.93	83.42	61.40
6.00	62.28	83.04	95.56	99.23	99.90	99.90	99.23	95.56	83.04	62.28
6.40	63.11	82.71	95.21	99.11	99.87	99.87	99.11	95.21	82.71	63.11
6.80	63.89	82.42	94.86	98.98	99.84	99.84	98.98	94.86	82.42	63.89
7.20	64.63	82.18	94.53	98.85	99.80	99.80	98.85	94.53	82.18	64.63
7.60	65.33	81.97	94.21	98.72	99.77	99.77	98.72	94.21	81.97	65.33
8.00	66.00	81.79	93.90	98.58	99.72	99.72	98.58	93.90	81.79	66.00
8.40	66.63	81.65	93.60	98.44	99.68	99.68	98.44	93.60	81.65	66.63
8.80	67.23	81.53	93.32	98.29	99.63	99.63	98.29	93.32	81.53	67.23
9.20	67.80	81.43	93.05	98.15	99.58	99.58	98.15	93.05	81.43	67.80
9.60	68.35	81.35	92.79	98.00	99.52	99.52	98.00	92.79	81.35	68.35
10.00	68.87	81.28	92.54	97.85	99.46	99.46	97.85	92.54	81.28	68.87
10.40	69.37	81.24	92.30	97.71	99.39	99.39	97.71	92.30	81.24	69.37
10.80	69.84	81.20	92.07	97.56	99.33	99.33	97.56	92.07	81.20	69.84
11.20	70.29	81.18	91.86	97.41	99.26	99.26	97.41	91.86	81.18	70.29
11.60	70.73	81.18	91.65	97.26	99.18	99.18	97.26	91.65	81.18	70.73
12.00	71.15	81.18	91.46	97.11	99.10	99.10	97.11	91.46	81.18	71.15
12.40	71.55	81.19	91.27	96.97	99.02	99.02	96.97	91.27	81.19	71.55
12.80	71.93	81.20	91.10	96.82	98.94	98.94	96.82	91.10	81.20	71.93
13.20	72.31	81.23	90.93	96.68	98.86	98.86	96.68	90.93	81.23	72.31
13.60	72.66	81.26	90.77	96.53	98.77	98.77	96.53	90.77	81.26	72.66

The dynamics is as expected: averaging of the temperature.

We also realize that this is a function of two variables! It can also be visualized by its graph:



Another experiment, with hot and cold ends:

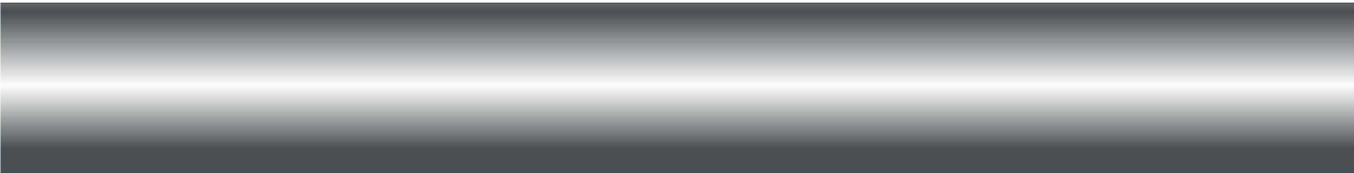


Once again, convergence toward the average temperature!

Exercise 6.1.5

Implement heat transfer in a circular arrangement.

The model is an approximation of a *metal rod*:



This is an insulated rod and might be heated at one end as above. The temperature then varies along its length even when the time is fixed. We need another variable!

We re-introduce the variables:

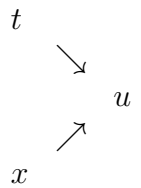
1. t is the time.
2. x is the location.
3. u is the temperature.

All three are just numbers.

How do they depend on each other?

There is no motion! Therefore, at every moment of time t one can be at every location x and vice versa. This means that t and x are independent of each other. Next, at every moment of time t and at every location x we can measure the temperature u . This means that u depends on t and x . So, we have the

following dependence diagram:



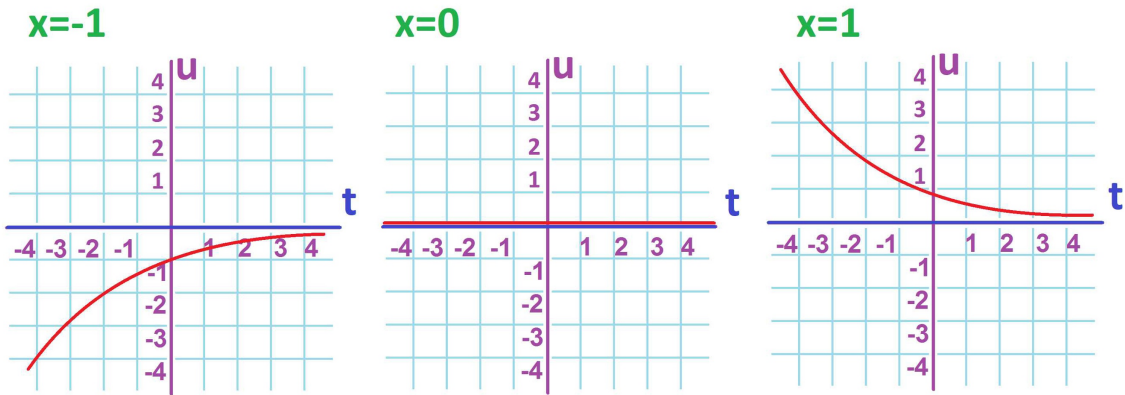
We simply have a *function of two variables*:

$$u = f(t, x).$$

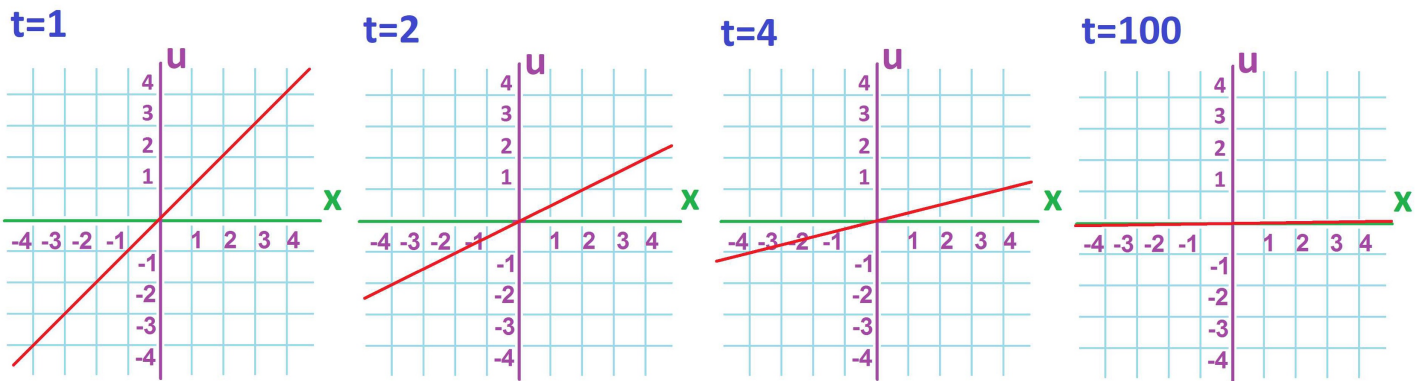
Before developing a model and a differential equation for this function, what is a reasonable solution that we can envision?

Let’s set up the initial conditions. Suppose the rod is cool – negative – at one end and hot – positive – at the other.

First we imagine that we sit next to a particular location on the rod and observe the dynamics of temperature over *time*. Judging by Newton’s Law, the temperature will gradually converge to a particular value over time. Suppose this value is 0:



Now we imagine that we “freeze” time and run along the rod observing how the temperature is changing over *space*. Suppose the temperature varies linearly from one end to the other. Judging by the Conservation of Energy Law, the temperature will gradually average itself with time. It is conceivable that the distribution of the temperature will remain linear:



We can try to imagine what the function f is like. The ODE has the exponential decay solutions:

$$u = r + Ce^{-kt}.$$

To get the general shape, we can try:

$$u = xe^{-t} \text{ or } u = x/t.$$

Let’s collect the data for this function: inputs and outputs. The inputs being independent of each other form an *array*, say x in rows and t in columns:

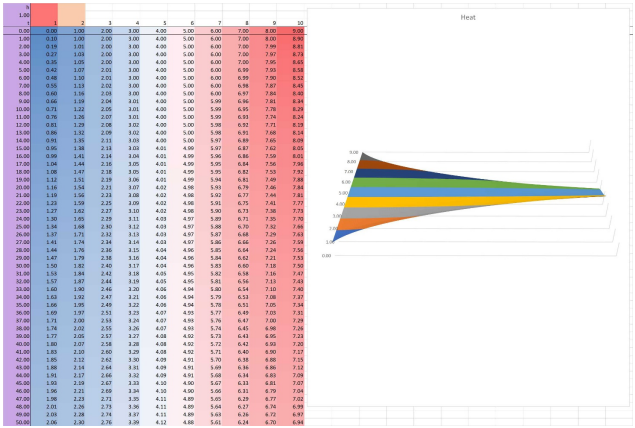
$t \backslash x$	0.0	1.5	2.0	2.5
0				
1				
2				

In other words, our function f has the (t, x) -plane or its subset for a domain.

Now the outputs. We just place them at each of the cells of the above table:

$t \backslash x$	0	1.5	2	2.5
0	1	2	0	
1	3	4	0	
2	0	0	−1	

Plotting the solutions to our difference equation matches this description:



This time, all but the objects at the ends follow the same formula that refers to its *two* neighbors:

$$v(t_{i+1}) = v(t_i) + k(u(t_i) - v(t_i)) \cdot \Delta t + k(w(t_i) - v(t_i)) \cdot \Delta t.$$

We rewrite the formula in the language of functions of two variables.

So, we have a function of t and x .

The time is as before. These are the nodes of our cell decomposition of the t -axis:

$$\dots, t_{i-1}, t_i, t_{i+1}, \dots,$$

with the increment:

$$\Delta t = t_{i+1} - t_i.$$

The space is similar. These are the nodes of our cell decomposition of the x -axis (the rod):

$$\dots, x_{n-1}, x_n, x_{n+1}, \dots$$

We are making a switch from three functions of one variables one function of two variables. In particular, we have this correspondence to the previous notation:

	$u(\cdot)$	$v(\cdot)$	$w(\cdot)$	
...	x_{n-1}	x_n	x_{n+1}	...
...	$u(\cdot, x_{n-1})$	$u(\cdot, x_n)$	$u(\cdot, x_{n+1})$...
...	—●—	—●—	—●—	...

The heat is exchanged between these “containers” through these “pipes”.

Then our formula,

$$v(t_{i+1}) = v(t_i) + k(u(t_i) - v(t_i)) \cdot \Delta t + k(w(t_i) - v(t_i)) \cdot \Delta t$$

becomes the following:

$$u(t_{i+1}, x_n) = u(t_i, x_n) + k(u(t_i, x_{n-1}) - u(t_i, x_n)) \cdot \Delta t + k(u(t_i, x_{n+1}) - u(t_i, x_n)) \cdot \Delta t.$$

The formula can be used for simulation as in the last example.

What is the meaning of the formula?

First, we notice the difference with respect to t :

$$\Delta_t u(t_i, x_n) = u(t_{i+1}, x_n) - u(t_i, x_n).$$

We add a subscript t to indicate the variable just as with partial derivatives!

Then, our formula becomes:

$$\Delta_t u(t_i, x_n) = k(u(t_i, x_{n-1}) - u(t_i, x_n)) \cdot \Delta t + k(u(t_i, x_{n+1}) - u(t_i, x_n)) \cdot \Delta t.$$

The two terms are the amounts of heat exchanged with the two neighbors.

Dividing by Δt simplifies things:

$$\frac{\Delta_t u(t_i, x_n)}{\Delta t} = k \left[u(t_i, x_{n-1}) - u(t_i, x_n) + u(t_i, x_{n+1}) - u(t_i, x_n) \right].$$

What is the meaning of the term in the right-hand side? Is this the sum of two differences?! No.

First, these are differences but not *differences* Δ . The temperature at the location is subtracted from that of a neighbor, no matter left or right, a smaller or larger position within the x -axis. This is the difference with respect to x :

$$\Delta_x u(t_i, x_n) = u(t_i, x_{n+1}) - u(t_i, x_n).$$

Right minus left! Let's find the differences with respect to space in the right-hand side of the equation (\cdot stands for t_i):

$$\begin{aligned} &u(\cdot, x_{n-1}) - u(\cdot, x_n) + u(\cdot, x_{n+1}) - u(\cdot, x_n) \\ &= -[u(\cdot, x_n - u(\cdot, x_{n-1}))] + [u(\cdot, x_{n+1}) - u(\cdot, x_n)] \\ &= -[\Delta_x u(\cdot, x_n)] + [\Delta_x u(\cdot, x_{n+1})]. \end{aligned}$$

This is the difference of the difference!

It makes sense. Suppose we have three adjacent objects with these temperatures:

4	7	9
---	---	---

The heat flows from right to left:

4	←	←	7	←	←	9
---	---	---	---	---	---	---

- Will the temperature go up or down in the left cell? Up! Why? Because the neighbor is warmer.
- Will the temperature go up or down in the right cell? Down! Why? Because the neighbor is cooler.
- Will the temperature go up or down in the *middle* cell? Down! Why? There are two neighbors, one cooler and the other warmer... They have opposite effects on this cell! So, why down?

↑	4	↑	↓	7	↓	↓	9	↓
---	---	---	---	---	---	---	---	---

This is why:

► Because 7 is closer to 9 than to 4.

The heat flows from right to left but the middle one loses more than it gives:

4

⇐

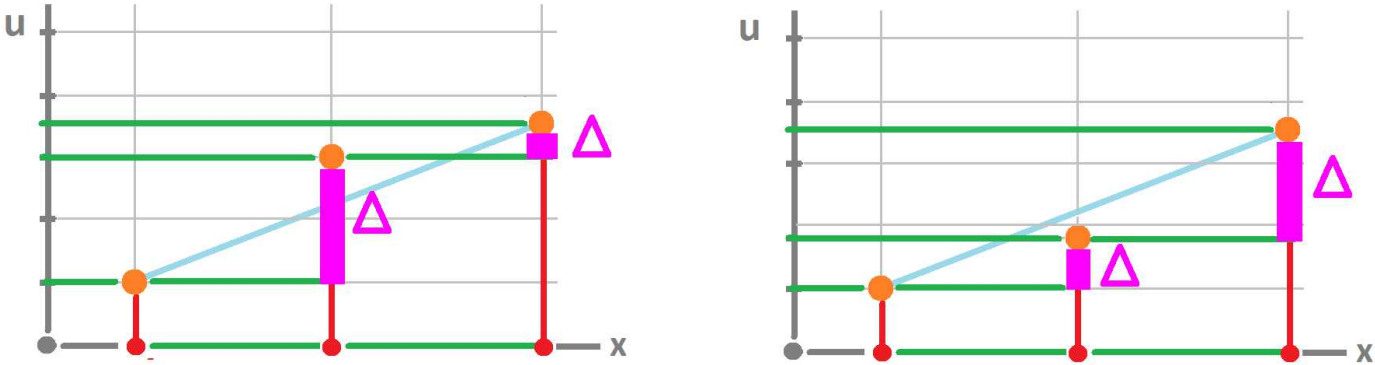
⇐ 7

←

← 9

Let’s take this apart. “Close” is about the differences. “Closer” is, therefore, about the differences of the differences!

This idea isn’t new; it’s all about *concavity* (and, eventually, the second derivative, [Chapter 2DC-4](#)) of the function:



Let’s find the second difference with respect to space in the right-hand side of our equation:

$$-\Delta_x u(\cdot, x_n) + \Delta_x u(\cdot, x_{n+1}) = \Delta_x \Delta_x u(\cdot, x_n) .$$

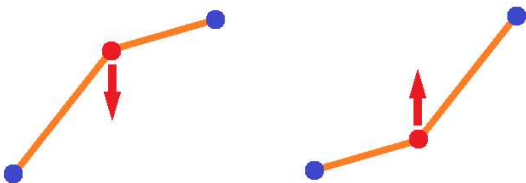
We have therefore:

$$\frac{\Delta_t u(t_i, x_n)}{\Delta t} = k \Delta_x \Delta_x u(t_i, x_n) .$$

Warning!

The simplification only works when k is constant.

So, the magnitude of the concavity determines by how much the value is pushed in the opposite direction:



The result is averaging.

This is the shortened version:

$$\frac{\Delta_t u}{\Delta t} = k \Delta_x^2 u$$

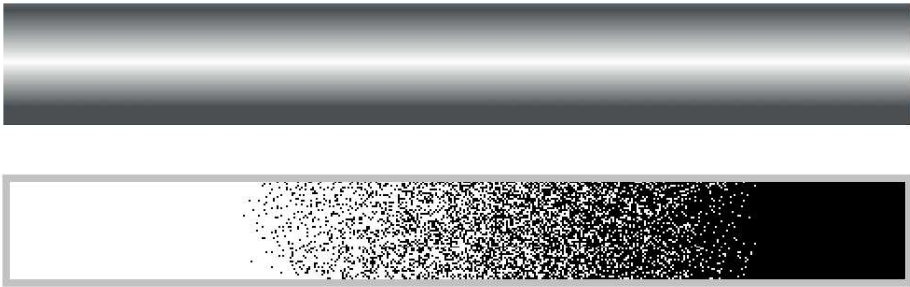
It is called the *heat difference equation*.

The reality is more complex:

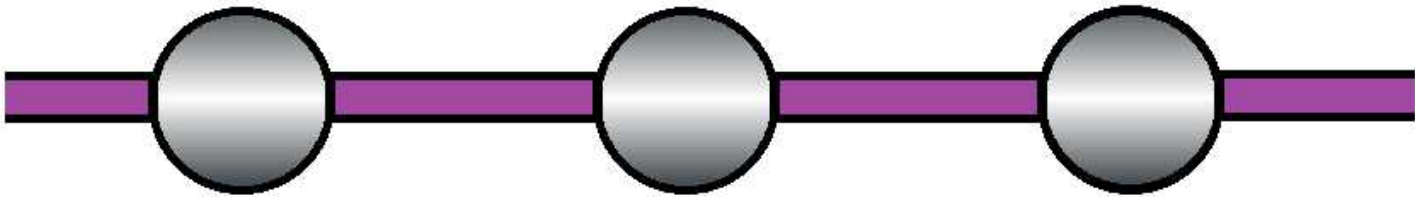
► The transfer rate might be different at different locations.

6.2. Heat transfer depends on permeability

We start in this section with heat transfer within a metal rod. It’s heat-transferring properties might be non-uniform:



The rod is seen in the following hypothetical way: as if it’s split into pieces. Furthermore, the heat is contained in a *string of containers* and each container exchanges the heat with its *two neighbors* through the pipes:



Here we have:

- x, y, z are some of the rooms.
- a, b are some of the pipes.



Of course, we replace the containers with nodes and pipes with edges.

A careful look reveals that to model heat transfer, we need to separately record the exchange of heat of each container with each of the adjacent containers.

The process we study obeys the following familiar law of physics. Newton’s Law of Cooling: The rate of cooling of an object is proportional to the difference between its temperature and the temperature an adjacent object.

Of course, cooling means heating when the temperature of the other object is higher.

Our initial assumption is that *all containers have equal size*.

It follows that the amount of heat in the containers is proportional to their respective temperatures. This is why we can understand the law of cooling as follows:

- *The rate of heat transfer between every two adjacent containers is proportional to the difference between the amounts of heat in them.*

The assumption of *conservation of energy* in container x gives us the following. The change of the amount of the heat in containers x over the time increment from t_i to t_{i+1} is equal to

$$u(t_{i+1}, x) - u(t_i, x) = - \left(\text{sum of the outflow } g \text{ through the pipes of } x \right).$$

The outflow gives the amount of flow across an edge (from the container to its neighbor) per unit of time. Specifically, the flow is positive at x if it is from left to right and the opposite for y ; then:

$$u(t_{i+1}, x) - u(t_i, x) = - \left(g(t_i, a) - g(t_i, b) \right) = g(t_i, a) - g(t_i, b).$$

Warning!

What is the magnitude of this rate? It cannot be so fast that the updated value of the heat of the container surpasses that of the other! In order to avoid this “overshoot”, the increment of the heat shouldn’t be more that half-difference from the heat of the other container.

Now, we need to express g in terms of u . The flow $g(t, a)$ through pipe a is proportional to the difference of the amounts of heat in a and the other container adjacent to a . So,

pipe a : $g(t_i, a) = -K(a)(u(t_i, x) - u(t_i, y))\Delta t$

pipe b : $g(t_i, b) = -K(b)(u(t_i, z) - u(t_i, x))\Delta t$

Here, $K(a) \geq 0$ represents the *permeability* of the pipe a at a given time over the same time period. These numbers $K(a)$ are simply the proportionality coefficients. They, therefore, produce a discrete 1-form. The benefit of this approach is that we won’t need to have separate formulas for the ends of the rod!

Exercise 6.2.1

Can K also depend on t ?

The result of the substitution is the following equation:

$$u(t_{i+1}, x) = u(t_i, x) + \left[-K(a)(u(t_i, x) - u(t_i, y)) \right] \Delta t - \left[-K(b)(u(t_i, z) - u(t_i, x)) \right] \Delta t$$

Let’s review the setup for discrete forms of one variable.

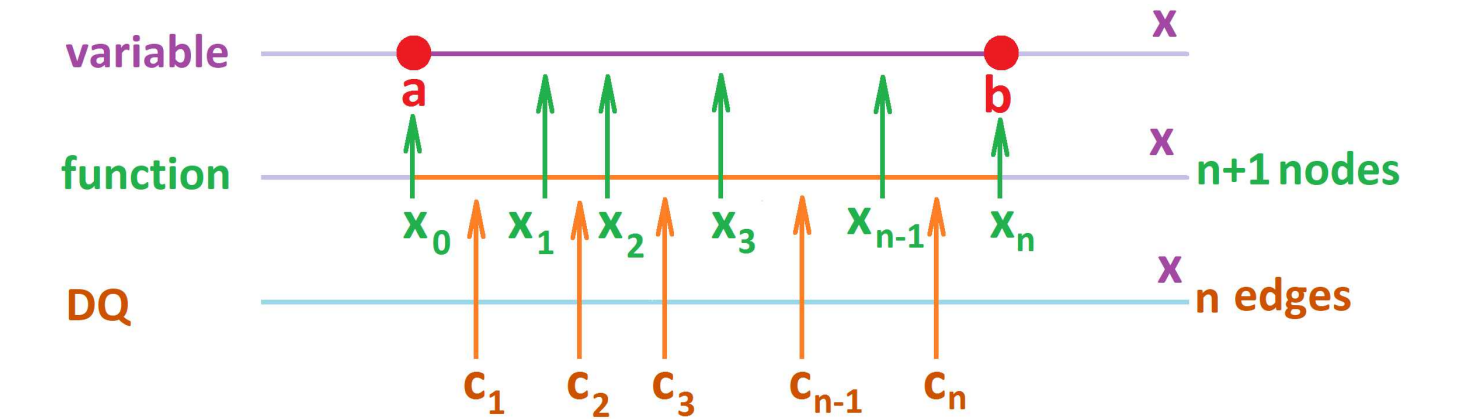
We have a *cell decomposition* of an interval $[a, b]$. We decompose it into n edges with the help of the nodes:

$$a = x_0, \ x_1, \ x_2, \ ..., \ x_{n-1}, \ x_n = b.$$

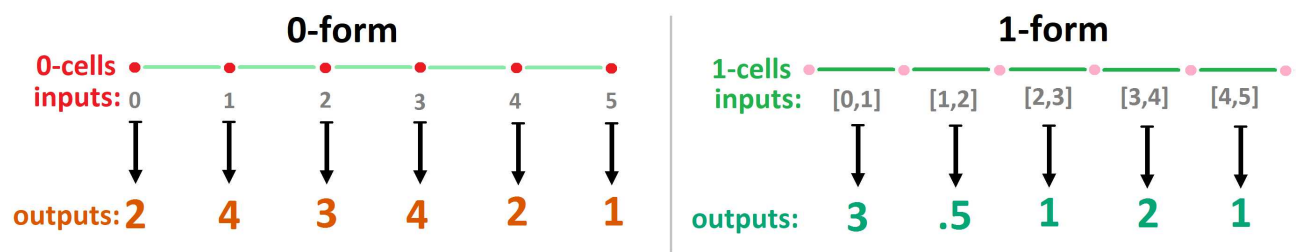
These are the edges:

$$c_1 = [x_0, x_1], \ c_2 = [x_1, x_2], \ ..., \ c_n = [x_{n-1}, x_n].$$

A function defined on the nodes is a 0-form and its difference or difference quotient is a 1-form:



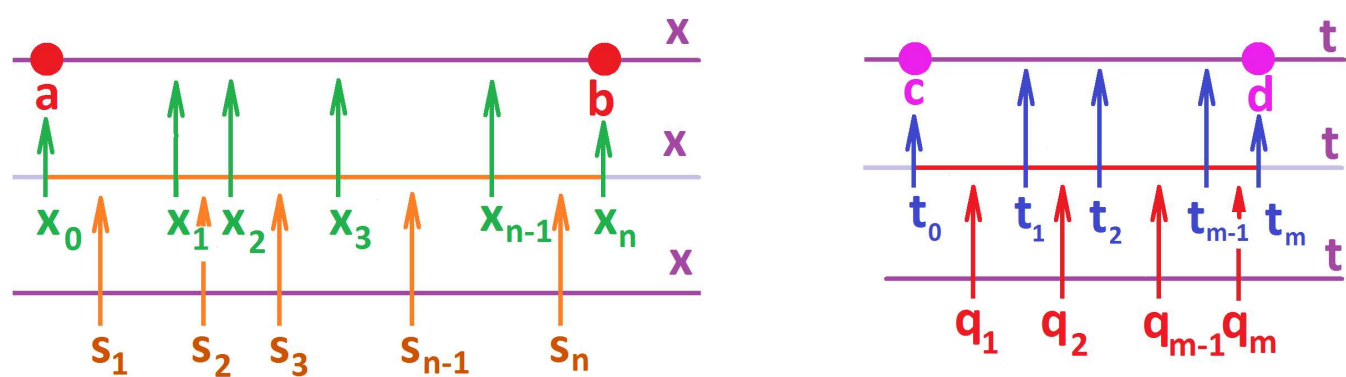
The 0-forms have nodes as inputs and the 1-forms have edges as inputs but the outputs are real numbers:



What makes this different from ODEs is that the functions will have two variables – one for location and one for time. The *amount of heat* $u = u(t, x)$ is simply a number assigned to each container x . This is

- 1. a discrete 0-form with respect to location x and
- 2. a discrete 0-form with respect to time t .

Let’s consider the setup for discrete forms of two variables.
We start with cell decompositions in the x - and the t -axes:



They develop in parallel:

- A cell decomposition of the interval $[a, b]$ in the x -axis consists of $n + 1$ nodes:

$$a = x_0, \, x_1, \, x_2, \, \dots, \, x_{n-1}, \, x_n = b,$$

and n edges:

$$s_1, \, s_2, \, \dots, \, s_{n-1}, \, s_n,$$

with the increments:

$$\Delta x_k = x_k - x_{k-1}, \, k = 1, 2, \dots, n.$$

- A cell decomposition the interval $[c, d]$ in the t -axis consists of $m + 1$ nodes:

$$c = t_0, \, t_1, \, t_2, \, \dots, \, t_{m-1}, \, t_m = d,$$

and m edges:

$$q_1, \, q_2, \, \dots, \, q_{m-1}, \, q_m,$$

with the increments:

$$\Delta t_i = t_i - t_{i-1}, \, i = 1, 2, \dots, m.$$

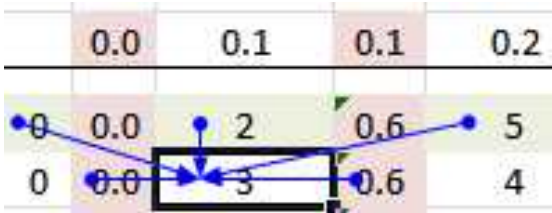
We use these decompositions of the intervals to construct a *cell decomposition* P of a rectangle $R = [a, b] \times [c, d]$ in the xt -plane. The lines $x = x_k$ and $t = t_i$ cut the rectangle $[a, b] \times [c, d]$ into smaller rectangles $[x_k, x_{k+1}] \times [t_i, t_{i+1}]$.

permeability k at that location at that time. The heat flow through the pipes is computed according to the formula discussed above and the new value of u is computed every time we move to the next row.

The spreadsheet formula is as follows:

=R[-1]C-RC[-1]*(R[-1]C-R[-1]C[-2])+RC[1]*(R[-1]C[2]-R[-1]C)

This is the dependence diagram of the formula:



Example 6.2.2: insulated rod

Suppose we put insulators to the ends of our rod and suppose they have initially a lower temperature than that of the rod. The areas adjacent to the ends quickly cools down. At this point we start our simulation. Heat transfer continues within the body of the rod with virtually no transfer through the insulators.

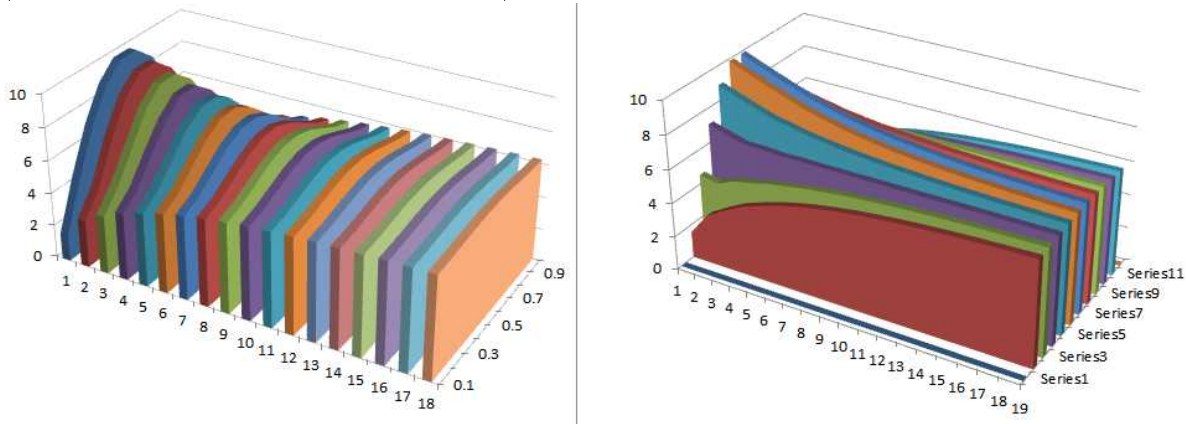
We set the initial value of u to go from 0 to 1 and then back to 0 at the other end. For example, we may have:

$$u(0, x) = 10 \sin(\pi x).$$

The permeability is zero at the ends and non-zero throughout the rest of the rod. For example, we may have:

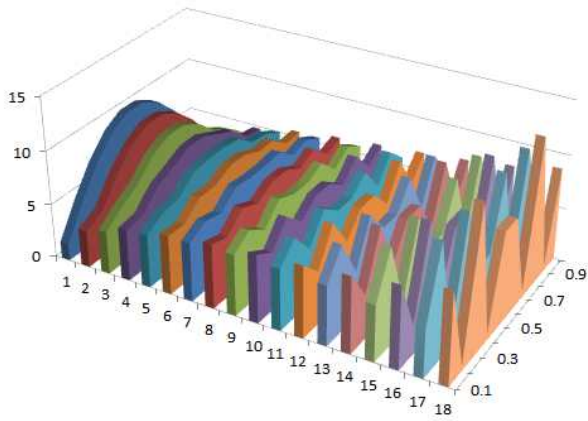
$$K(0) = K(1) = 0 \text{ and } K(x) = .45 \text{ for } 0 < x < 1.$$

This is the result of our simulation is the collection of graphs of the function $u(t, \cdot)$ of one variable for each t (the graph of u is of course a surface):



Above we see either the distribution of heat throughout the rod at every moment of time or the dynamics of the temperature at every location. Conclusion: the temperature evens out! Note that the total amount of heat in the rod remains the same (given under “total” in the spreadsheet).

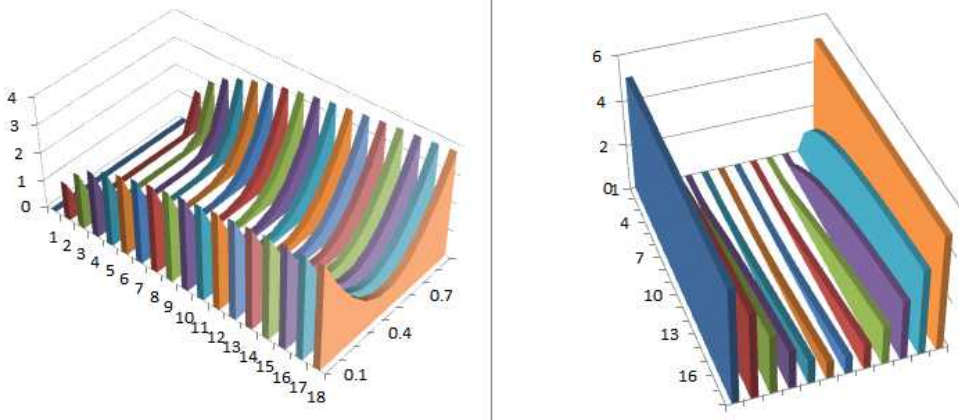
As a warning, if we choose our coefficient of permeability larger than 1/2, the simulation fails quickly:



Example 6.2.3: heated rod

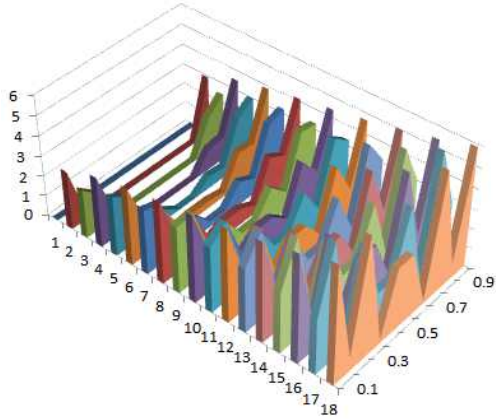
Suppose our rod is inserted to fix a gap in a long pipe with warmer temperature. Once in, the areas adjacent to the ends of the rod start to exchange heat with the pipe. However, the effect on the pipe's temperature is negligible. At this point we start our simulation.

We set the initial value of u to be 0 throughout the rod while the outside temperature is 5. The permeability is .45 throughout the rod. This is the result of our simulation:



The temperature evens out!

If we choose our coefficient of permeability larger than 1/2, the simulation fails quickly:



Exercise 6.2.4

What kind of curve do we see at the edge ($x = 0$) of this surface?

Example 6.2.5: simplified

A simplified version makes k independent of time and puts it at the top row. This is the dependence:

t	x=	1	2	3
0.00		0.00	0.00	0.00
1.00		0.10	0.01	0.00
2.00		0.19	0.03	0.01

The result is on the left:

t	x=	0.0	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0
0.00		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.00		0.10	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
2.00		0.19	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
3.00		0.27	0.05	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
4.00		0.35	0.08	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
5.00		0.43	0.12	0.06	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
6.00		0.49	0.15	0.07	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
7.00		0.56	0.19	0.09	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
8.00		0.62	0.22	0.11	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09
9.00		0.68	0.26	0.13	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
10.00		0.74	0.30	0.15	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12
11.00		0.80	0.34	0.17	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13
12.00		0.85	0.37	0.20	0.15	0.14	0.14	0.14	0.14	0.14	0.14	0.14
13.00		0.90	0.41	0.22	0.16	0.15	0.15	0.15	0.15	0.15	0.15	0.15
14.00		0.95	0.45	0.24	0.18	0.16	0.16	0.16	0.16	0.16	0.16	0.16
15.00		1.00	0.49	0.27	0.19	0.17	0.17	0.17	0.17	0.17	0.17	0.17
16.00		1.05	0.53	0.29	0.21	0.18	0.18	0.18	0.18	0.18	0.18	0.18
17.00		1.10	0.56	0.32	0.23	0.19	0.19	0.19	0.19	0.19	0.19	0.19
18.00		1.15	0.60	0.34	0.24	0.20	0.20	0.20	0.20	0.20	0.20	0.20
19.00		1.19	0.64	0.37	0.26	0.21	0.21	0.21	0.21	0.21	0.21	0.21
20.00		1.24	0.67	0.40	0.28	0.22	0.22	0.22	0.22	0.22	0.22	0.22

t	x=	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0
0.00		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.00		0.10	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
2.00		0.19	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
3.00		0.27	0.05	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
4.00		0.35	0.08	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
5.00		0.43	0.12	0.06	0.05	0.05	0.05	0.05	0.05	0.05	0.05
6.00		0.49	0.15	0.07	0.06	0.06	0.06	0.06	0.06	0.06	0.06
7.00		0.56	0.19	0.09	0.08	0.08	0.08	0.08	0.08	0.08	0.08
8.00		0.62	0.22	0.11	0.09	0.09	0.09	0.09	0.09	0.09	0.09
9.00		0.68	0.26	0.13	0.10	0.10	0.10	0.10	0.10	0.10	0.10
10.00		0.74	0.30	0.15	0.12	0.12	0.12	0.12	0.12	0.12	0.12
11.00		0.79	0.31	0.16	0.13	0.13	0.13	0.13	0.13	0.13	0.13
12.00		0.84	0.34	0.17	0.14	0.14	0.14	0.14	0.14	0.14	0.14
13.00		0.89	0.37	0.19	0.15	0.15	0.15	0.15	0.15	0.15	0.15
14.00		0.94	0.40	0.20	0.16	0.16	0.16	0.16	0.16	0.16	0.16
15.00		0.99	0.42	0.21	0.17	0.17	0.17	0.17	0.17	0.17	0.17
16.00		1.03	0.45	0.22	0.18	0.18	0.18	0.18	0.18	0.18	0.18
17.00		1.07	0.47	0.23	0.19	0.19	0.19	0.19	0.19	0.19	0.19
18.00		1.11	0.50	0.24	0.20	0.20	0.20	0.20	0.20	0.20	0.20
19.00		1.15	0.52	0.25	0.21	0.21	0.21	0.21	0.21	0.21	0.21
20.00		1.19	0.54	0.26	0.22	0.22	0.22	0.22	0.22	0.22	0.22

Exercise 6.2.6

What happened in the second spreadsheet?

Example 6.2.7: heated rod

What will happen to the rod heated at the ends when the *permeability varies*? The initial temperature is still 0 while the outside temperature is 5. The permeability will remain .45 throughout the half of the rod and .1 on the left. This is the result of our simulation:

The temperature evens out again but, predictably, the left end is falling behind!

Warning!

The rod is its own universe.

We will now interpret the differences we see in our recursive formula:

$$u(t_{i+1}, x_k) = u(t_i, x_k) - K(s_k)(u(t_i, x_k) - u(t_i, x_{k-1}))\Delta t + K(s_{k+1})(u(t_i, x_{k+1}) - u(t_i, x_k))\Delta t.$$

If a function $y = f(x)$ is defined at the nodes x_k , $k = 0, 1, 2, \dots, n$, the *difference* of f is defined at the edges by:

$$\Delta f(c_k) = f(x_{k+1}) - f(x_k)$$

for each $k = 1, 2, \dots, n$.

The partial *differences of u with respect to x and t* are defined at these edges by the following:

$$\Delta_x u(t_i, s_k) = u(t_i, x_{k+1}) - u(t_i, x_k) \text{ and } \Delta_t u(q_i, x_k) = u(t_{i+1}, x_k) - u(t_i, x_k)$$

We take our recursive formula and rewrite it in terms of the difference quotient:

$$\frac{\Delta u}{\Delta t}(q_i, x_k) = K(s_{k+1})\Delta_x u(t_i, s_{k+1}) - K(s_k)\Delta_x u(t_i, s_k)$$

What is the meaning of the right-hand side?

6.3. Heat transfer is caused by temperature differences

Example 6.3.1: large tea cup

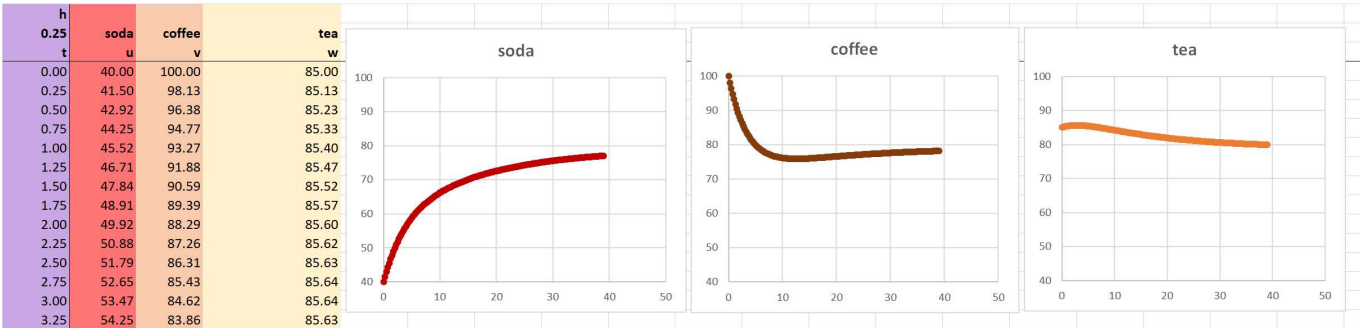
What if the tea cup it 3 times larger than the other two?



We adjust the formula by dividing by 3 the amount that is added to w (tea):

$$=R[-1]C-0.1\cdot R2C\cdot (R[-1]C-R[-1]C[-1])\cdot R2C1-0.1\cdot R2C[1]\cdot (R[-1]C-R[-1]C[1])\cdot R2C1/3$$

The idea is that is the effect of the difference of the temperatures on the tea is one third:



The pattern of averaging continues but the average has changed:

$$\frac{40 + 100 + 85 \cdot 3}{5} = 79.$$

A “quick-and-dirty” fix has produced the expected result!

We continue with a more general model for heat transfer in a rod. This time, the containers’ sizes may vary and so do the lengths of the intervals of time.

What matters now is the juxtaposition:

- heat vs. temperature,
- weight vs. density.

The latter is the average amount of the former:

$$\text{temperature} = \frac{\text{heat}}{\text{size}}$$

and

density = $\frac{\text{weight}}{\text{size}}$.

The sizes are nothing but a sequence of positive numbers assigned to each node x_k . Let’s call them Δs_k . (The latter formula matches the one in [Chapter 3IC-3](#).)

If u is the temperature, then the amount of heat in the room x is:

$u(\cdot, x_k) \cdot \Delta s_k.$

It is heat that is being exchanged. But it is the difference of the temperatures that drives the exchange. Our recursive formula becomes:

$u(t_{i+1}, x_k) \cdot \Delta s_k = u(t_i, x_k) \cdot \Delta s_k + K(s_{k+1})\Delta_x u(t_i, s_{k+1})\Delta t - K(s_k)\Delta_x u(t_i, s_k)\Delta t.$

Warning!

The size Δs_k varies from location to location.

The new recursive formula for the temperature is:

$u(t_{i+1}, x_k) = u(t_i, x_k) + [K(s_{k+1})\Delta_x u(t_i, s_{k+1}) - K(s_k)\Delta_x u(t_i, s_k)] \Delta t / \Delta s_k$

We use it for simulations.

Example 6.3.2: rod

We place the corresponding value of Δs at the top of each column:

h																
1.00	k= 0.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0				
	Delta s	1.0	1.0	1.0	1.0	1.0	10.0	1.0	1.0	1.0	1.0					
t	x=	1	2	3	4	5	6	7	8	9	10		heat	temp		
0.00		1.00	1.00	1.00	1.00	1.00	2.00	3.00	3.00	3.00	3.00		37.00	1.95		
1.00		1.00	1.00	1.00	1.00	1.10	2.00	2.90	3.00	3.00	3.00		37.00	1.95		
2.00		1.00	1.00	1.00	1.01	1.18	2.00	2.82	2.99	3.00	3.00		37.00	1.95		
3.00		1.00	1.00	1.00	1.03	1.25	2.00	2.76	2.97	3.00	3.00		37.00	1.95		
4.00		1.00	1.00	1.00	1.05	1.30	2.00	2.70	2.95	3.00	3.00		37.00	1.95		
5.00		1.00	1.00	1.01	1.07	1.34	2.00	2.66	2.93	2.99	3.00		37.00	1.95		
6.00		1.00	1.00	1.01	1.09	1.38	2.00	2.62	2.91	2.99	3.00		37.00	1.95		
7.00		1.00	1.00	1.02	1.11	1.41	2.00	2.59	2.89	2.98	3.00		37.00	1.95		
8.00		1.00	1.00	1.03	1.13	1.44	2.00	2.56	2.87	2.97	3.00		37.00	1.95		
9.00		1.00	1.01	1.03	1.15	1.47	2.00	2.53	2.85	2.97	2.99		37.00	1.95		
10.00		1.00	1.01	1.04	1.17	1.49	2.00	2.51	2.83	2.96	2.99		37.00	1.95		
11.00		1.00	1.01	1.05	1.19	1.51	2.00	2.49	2.81	2.95	2.99		37.00	1.95		
12.00		1.00	1.01	1.06	1.21	1.53	2.00	2.47	2.79	2.94	2.98		37.00	1.95		
13.00		1.00	1.02	1.07	1.23	1.54	2.00	2.46	2.77	2.93	2.98		37.00	1.95		
14.00		1.01	1.02	1.08	1.24	1.56	2.00	2.44	2.76	2.92	2.97		37.00	1.95		
15.00		1.01	1.03	1.09	1.26	1.57	2.00	2.43	2.74	2.91	2.97		37.00	1.95		
16.00		1.01	1.03	1.10	1.27	1.58	2.00	2.42	2.73	2.90	2.96		37.00	1.95		
17.00		1.01	1.04	1.11	1.29	1.59	2.00	2.41	2.71	2.89	2.96		37.00	1.95		
18.00		1.01	1.04	1.12	1.30	1.60	2.00	2.40	2.70	2.88	2.95		37.00	1.95		
19.00		1.02	1.05	1.13	1.31	1.61	2.00	2.39	2.69	2.87	2.94		37.00	1.95		
20.00		1.02	1.05	1.14	1.32	1.62	2.00	2.38	2.68	2.86	2.93		37.00	1.95		
21.00		1.02	1.06	1.15	1.33	1.63	2.00	2.37	2.66	2.85	2.93		37.00	1.95		

The formulas is as follows:

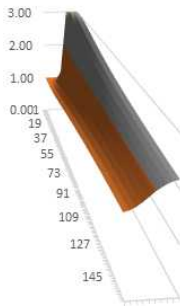
=R[-1]C-0.1*R2C*(R[-1]C-R[-1]C[-1])*R2C1/R3C-0.1*R2C[1]*(R[-1]C-R[-1]C[1])*R2C1/R3C

This is the new dependence:

h				
1.00	k= 0.0	1.0	1.0	
Delta s	1.0	1.0	1.0	
t	x=	1	2	3
0.00		1.00	1.00	1.00
1.00		1.00	1.00	1.00
2.00		1.00	1.00	1.00
3.00		1.00	1.00	1.00

We also compute the total heat and the average temperature at every moment of time; the last two columns. The conservation of energy is confirmed!

The averaging looks a bit different:



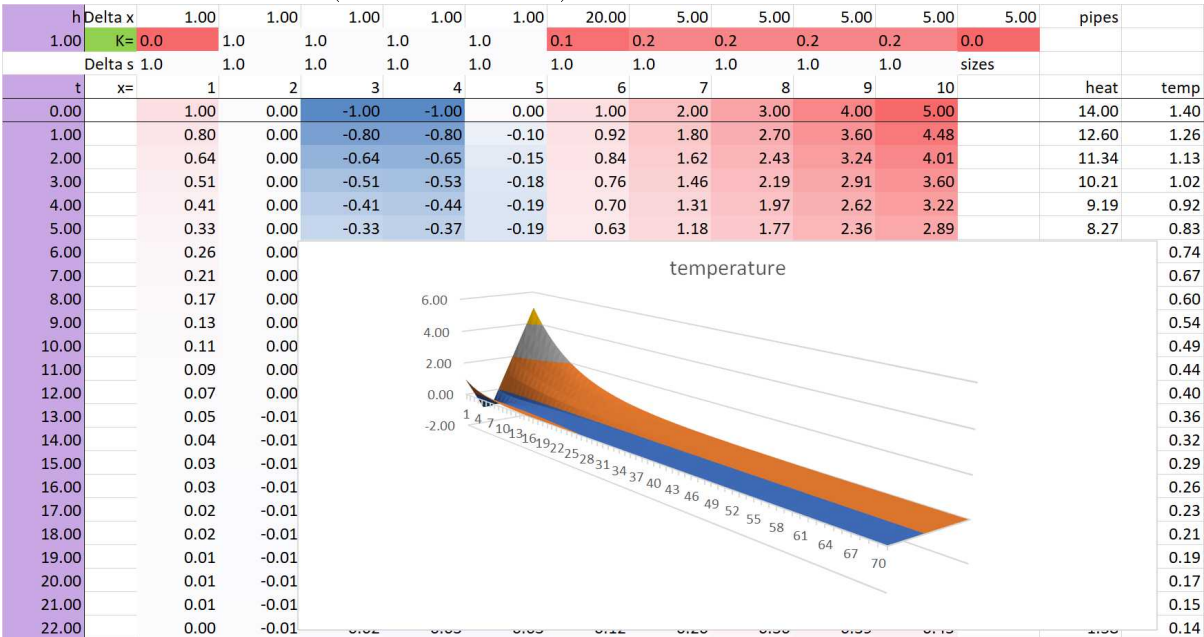
Example 6.3.3: exchange with outside

Now, suppose the rod is not its own universe this time and interacts with the outside. Then, in addition to the exchange between its cells, there is exchange with the air at room temperature. The latter is executed just as in the original Newton’s Law model but for each container independently!

We just add an extra term to the formula (0 room temperature):

-0.1*R[-1]C

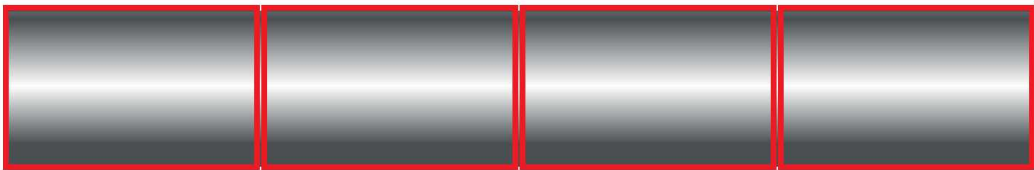
Of course, there is no conservation of energy this time. We can see how the temperature stabilizes toward the room temperature (the steady state):



This observation confirms the original Newton’s Law model.

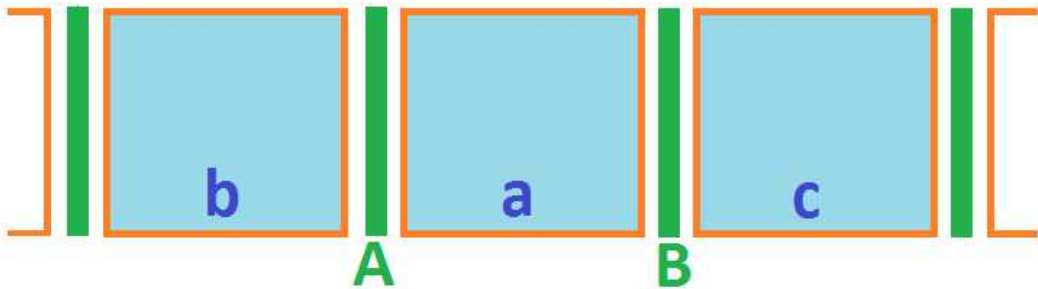
An alternative point of view on the rod and the heat exchange is as follows.

The rod is seen as split into pieces as if they are separate pieces: the heat is contained in a *row of rooms* and each room exchanges the material with its *two neighbors* through its walls:

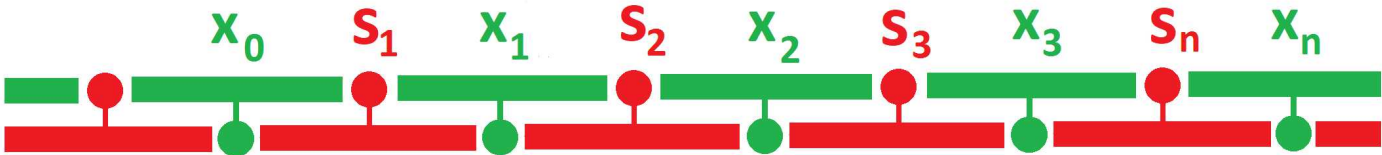


Below we have:

- $a = AB$ is one of the rooms.
- b, c are the two adjacent rooms, left and right.
- A, B are the walls of a , left and right.



Of course, we replace the rooms with edges and walls with nodes:

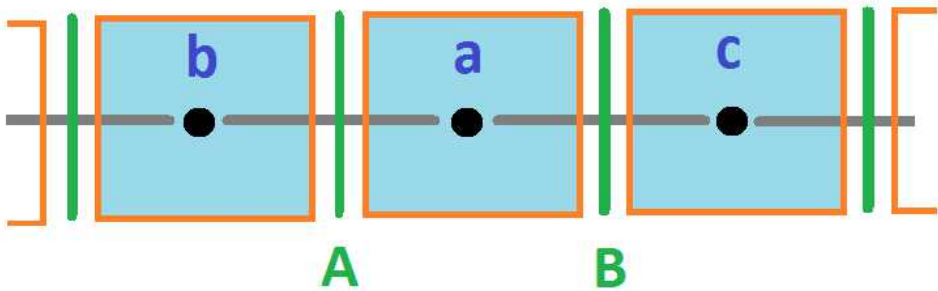


The interpretation is different from the one above; instead of rooms and walls between them, we spoke of containers and pipes between them.

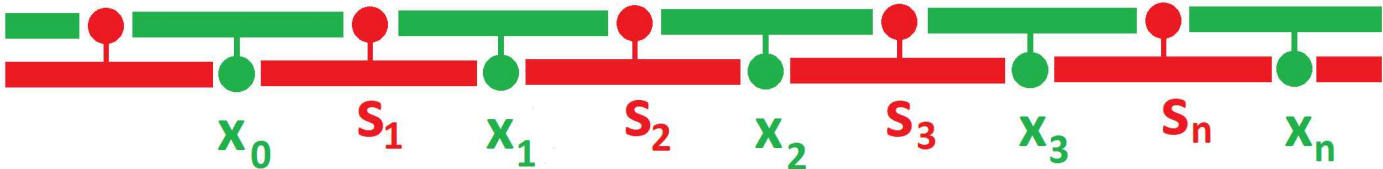
Recall:

- a is one of the containers.
- b, c are the two adjacent containers, left and right.
- A, B are the pipes that start at a , left and right.

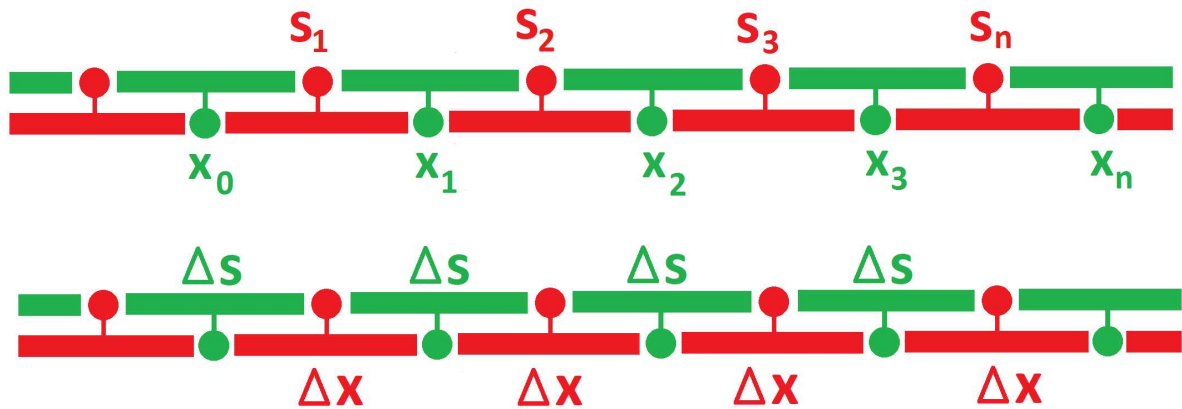
The correspondence between the two is illustrated below:



Of course, we replace the containers with nodes and the pipes with edges:



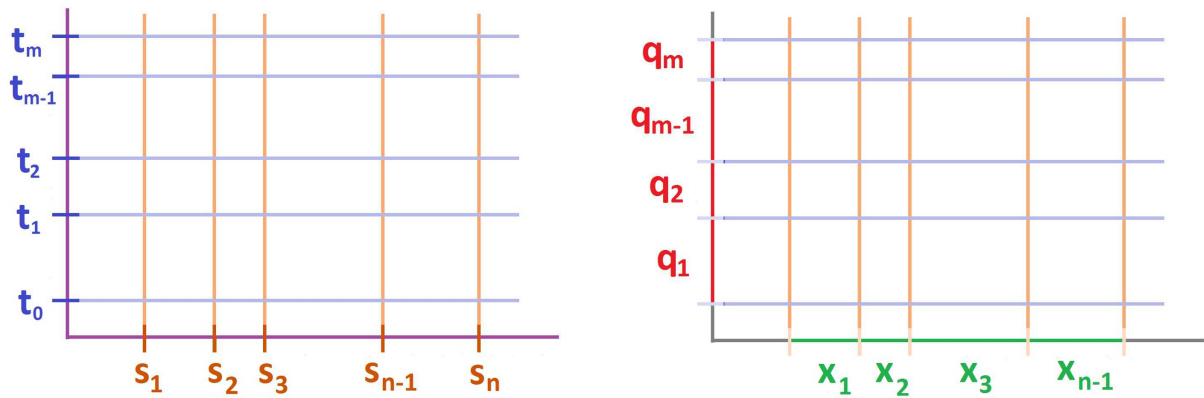
This is *duality*: Two identical steps and we are back where we started.
Let's review. We use the following match between the cells:



The lengths of the cells are indicated in the second row:

$$\Delta x_k = x_{k+1} - x_k \text{ and } \Delta s_k = s_{k+1} - s_k .$$

The former ones haven't appeared yet.
Under the *star operator*, the 0-cells of the former correspond to the 1-cells of the latter and vice versa.
The cell decompositions of our rectangle dual to each other are shown below:



We turn to our difference equation:

$$u(t_{i+1}, x_k) = u(t_i, x_k) + \left[K(s_{k+1}) \Delta_x u(t_i, s_{k+1}) - K(s_k) \Delta_x u(t_i, s_k) \right] \Delta t / \Delta s_k ,$$

or,

$$\frac{\Delta u}{\Delta t}(q_i, x_k) = \frac{K(s_{k+1}) \Delta_x u(t_i, s_{k+1}) - K(s_k) \Delta_x u(t_i, s_k)}{\Delta s_k} ,$$

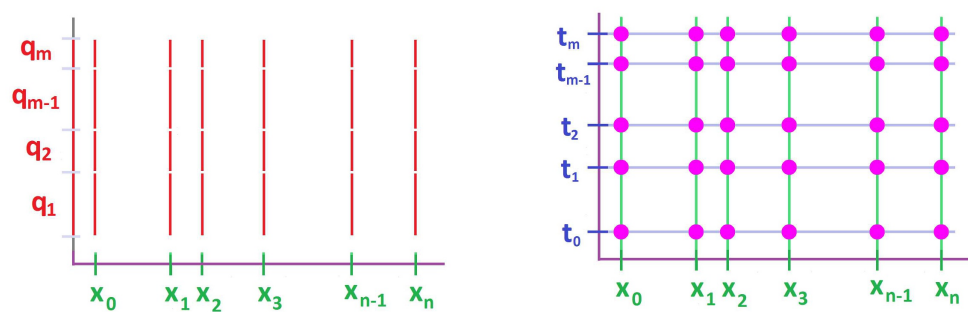
or,

$$\frac{\Delta u}{\Delta t}(q_i, x_k) = \frac{\Delta [K \Delta_x u]}{\Delta s_k}(t_i, x_k) .$$

When K is constant, we have an abbreviated form:

$$\frac{\Delta u}{\Delta t}(q_i, x_k) = K \frac{\Delta(\Delta_x u)}{\Delta s_k}(t_i, x_k)$$

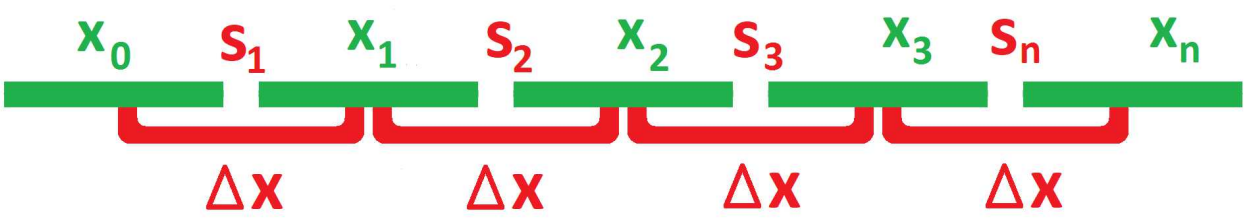
This is where the discrete forms in the left- and right-hand terms of the equation reside, respectively:



6.4. Heat transfer depends on the geometry

We start from scratch.

This is what our system of containers and pipes looks like:



In this section, we will pursue a new interpretation of the permeability K .

Let’s review the setup.

We decompose the segment $[a, b]$ into $n - 1$ intervals:

$$a = x_1, \, x_2, \, x_3, \, \dots, \, x_{n-1}, \, x_n = b.$$

Then the increments are:

$$\Delta x_k = x_{k+1} - x_k.$$

Suppose the *amount of flow* – left to right – of heat (or material) along the pipe

$$s_k = [x_{k-1}, x_k]$$

during a period of time

$$q_j = [t_{j-1}, t_j]$$

is denoted by:

$$g = g(q_j, s_k).$$

The *conservation of energy* (or material) gives us the following:

- The change of the amount of the heat in container x over a given interval of time is equal to the sum – accounting for the directions – of the flow along the two pipes that leave x .

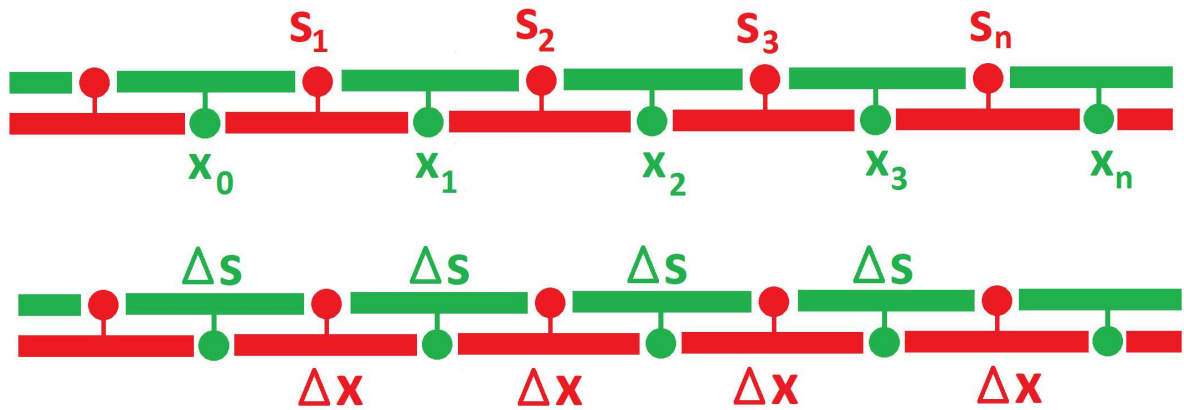
The statement takes the following algebraic form:

$$\Delta_t u(q_i, x_k) = g(q_i, s_{k-1}) - g(q_i, s_k)$$

Next, we separately record the exchange of heat between each pair of adjacent containers according to the *Newton’s Law of Cooling*: The rate of cooling of an object is proportional to the difference between its temperature and the temperature of the adjacent object.

In the last section, in addition to the amount of heat (or material), we included the temperature (or the density of the material, respectively) in our analysis.

The heat resides and moves within the *dual decomposition* in the x -axis:



We have utilized the idea that the length of the interval $x_k = [s_{k-1}, s_k]$ in the decomposition corresponds to the *size of the container* x_k :

$$\Delta s_k = s_k - s_{k-1} .$$

The new idea is that the length of the interval $s_k = [x_{k-1}, x_k]$ in the decomposition corresponds to the *size of the pipe* s_k :

$$\Delta x_k = x_k - x_{k-1} .$$

So, the law of cooling takes the following form:

- 1. The flow of heat (during a period of time) along a pipe is proportional to the difference of the temperature in the two containers at the two ends of the pipe (in the beginning of the period of time).

Furthermore, we have assumed the following:

- 2. The flow is proportional to the length of the period of time.

In the last section, we also made one more assumption:

- 3. The effect of the flow on the temperature is inversely proportional to the length of the compartment.

We make a *new assumption*:

- 4. The flow is inversely proportional to the length of the pipe.

We deal with these four quantities:

1. $u(t_i, x_k) - u(t_i, x_{k-1}) = \Delta_x u(t_i, s_k)$
2. Δt_i
3. $\frac{1}{\Delta s_k}$
4. $\frac{1}{\Delta x_k}$

In order to take into account the new item (#4), simply reinterpret the permeability:

$$K(s_k) = m(s_k) \frac{1}{\Delta x_k} ,$$

where $m(s_k)$ reflects the carrying capacity of the pipe.

The *geometry* of the system of pipes now comes into play.

Example 6.4.1

The lengths of pipes and of the compartments interact to produce a rigid structure:

We substitute these two into our formula and the result is the following:

$$\frac{\Delta u}{\Delta t}(q_i, x_k) = \frac{m(s_{k+1})\frac{1}{\Delta x_{k+1}}\Delta_x u(t_i, s_{k+1}) - m(s_k)\frac{1}{\Delta x_k}\Delta_x u(t_i, s_k)}{\Delta s_k}.$$

We can use it for simulation.

Example 6.4.2: rod

We place the value of Δx at the top of each column:

h	Delta x	1.00	1.00	1.00	1.00	1.00	20.00	5.00	5.00	5.00	5.00	5.00	pipes	
1.00	k= 0.0	1.0	1.0	1.0	1.0	1.0	0.1	0.2	0.2	0.2	0.2	0.0		
	Delta s 1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	sizes	
t	x=	1	2	3	4	5	6	7	8	9	10		heat	temp
0.00		1.00	1.00	1.00	1.00	1.00	3.00	3.00	3.00	3.00	3.00		20.00	2.00
1.00		1.00	1.00	1.00	1.00	1.01	2.99	3.00	3.00	3.00	3.00		20.00	2.00
2.00		1.00	1.00	1.00	1.00	1.02	2.98	3.00	3.00	3.00	3.00		20.00	2.00
3.00		1.00	1.00	1.00	1.00	1.03	2.97	3.00	3.00	3.00	3.00		20.00	2.00
4.00		1.00	1.00	1.00	1.00	1.03	2.96	3.00	3.00	3.00	3.00		20.00	2.00
5.00		1.00	1.00	1.00	1.01	1.04	2.95	3.00	3.00	3.00	3.00		20.00	2.00
6.00		1.00	1.00	1.00	1.01	1.05	2.94	3.00	3.00	3.00	3.00		20.00	2.00
7.00		1.00	1.00	1.00	1.01	1.05	2.94	3.00	3.00	3.00	3.00		20.00	2.00
8.00		1.00	1.00	1.00	1.02	1.06	2.93	3.00	3.00	3.00	3.00		20.00	2.00
9.00		1.00	1.00	1.00	1.02	1.06	2.92	2.99	3.00	3.00	3.00		20.00	2.00
10.00		1.00	1.00	1.01	1.02	1.07	2.91	2.99	3.00	3.00	3.00		20.00	2.00
11.00		1.00	1.00	1.01	1.02	1.07	2.90	2.99	3.00	3.00	3.00		20.00	2.00
12.00		1.00	1.00	1.01	1.03	1.08	2.90	2.99	3.00	3.00	3.00		20.00	2.00
13.00		1.00	1.00	1.01	1.03	1.08	2.89	2.99	3.00	3.00	3.00		20.00	2.00

We simply change the formulas for the permeability as follows:

=1/R[-1]C

This is the new dependence:

h	Delta x	1.00	1.00	1.00	1.00
1.00	k= 0.0	1.0	1.0	1.0	1.0
	Delta s 1.0	1.0	1.0	1.0	1.0
t	x=	1	2	3	4
0.00		1.00	1.00	1.00	1.00
1.00		1.00	1.00	1.00	1.00
2.00		1.00	1.00	1.00	1.01

This way, the rest of formulas remain the same.

The averaging looks different in the two halves of the rod:

We now interpret our recursive formula in terms of the second difference quotient.

We decompose the segment into $n - 1$ intervals by giving nodes to the edges of the last decomposition with the same names:

$$p = s_1, \, s_2, \, s_3, \, \dots, \, s_{n-1}, \, s_n = q.$$

Then the increments are:

$$\Delta s_k = s_{k+1} - s_k.$$

Now, what are the nodes corresponding to the edges of this new decomposition? The nodes of the last decomposition of course! Indeed, we have:

$$x_1 \text{ in } [s_1, s_2], \, x_2 \text{ in } [s_2, s_3], \, \dots, \, x_{n-1} \text{ in } [s_{n-1}, s_n].$$

We apply the same constructions to this decomposition to the function $g = \frac{\Delta f}{\Delta x}$. The difference function of g is defined at the edges of the new decomposition by:

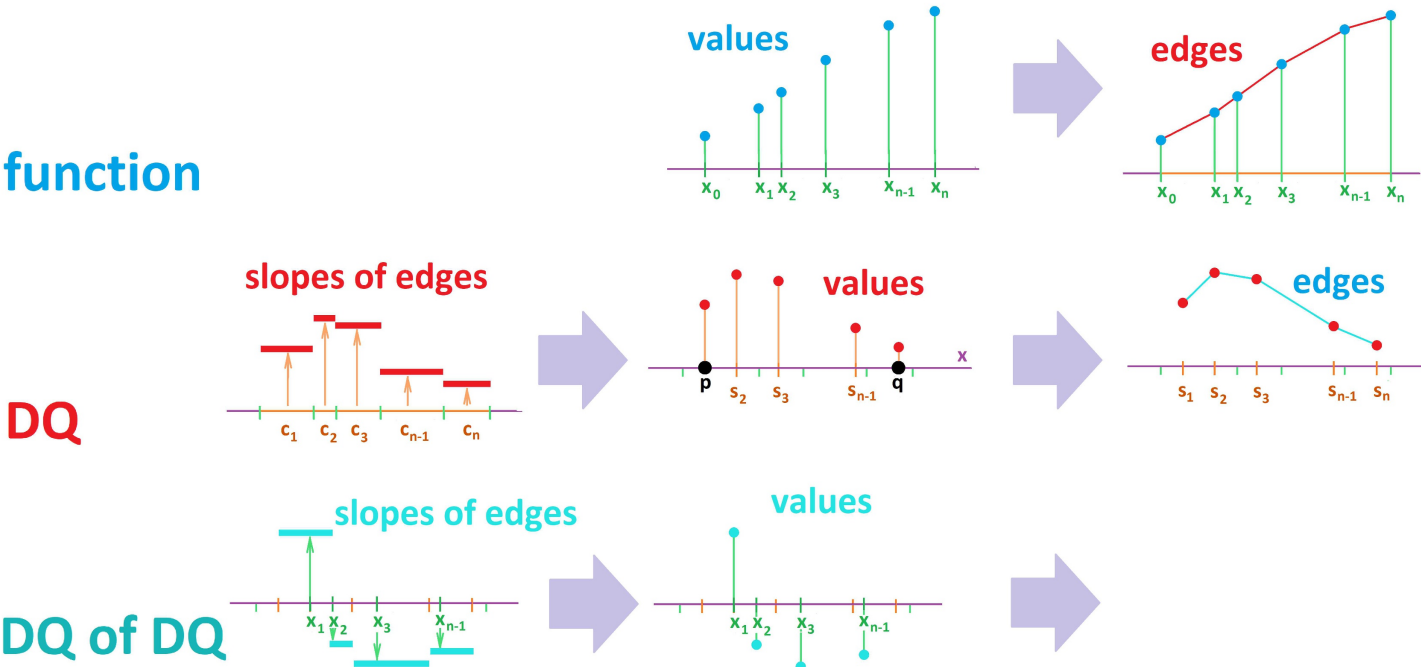
$$\Delta g(x_k) = g(s_{k+1}) - g(s_k)$$

for each $k = 1, 2, \dots, n - 1$. The difference quotient function of g is defined at the edges of the new decomposition by:

$$\frac{\Delta g}{\Delta x}(x_k) = \frac{g(s_{k+1}) - g(s_k)}{s_{k+1} - s_k}$$

for each $k = 1, 2, \dots, n - 1$.

The contrsruction is outlines in the middle row below:



The rest is in the bottom row.

The *second difference* of a function f is defined to be the difference of the difference, i.e., it is defined at the nodes of the decomposition (and denoted) as follows:

$$\Delta^2 f(x_k) = \Delta f(s_{k+1}) - \Delta f(s_k)$$

for each $k = 1, 2, \dots, n - 1$.

The *second difference quotient* of a function f is defined to be the difference quotient of the difference quotient, i.e., it is defined at the nodes of the decomposition (and denoted) as follows:

$$\frac{\Delta^2 f}{\Delta x^2}(x_k) = \frac{\frac{\Delta f}{\Delta x}(s_{k+1}) - \frac{\Delta f}{\Delta x}(s_k)}{s_{k+1} - s_k}$$

for each $k = 1, 2, \dots, n - 1$.

The second difference quotient is defined on the original decomposition and, therefore, is ready to be differentiated again!

Back to the heat equation.

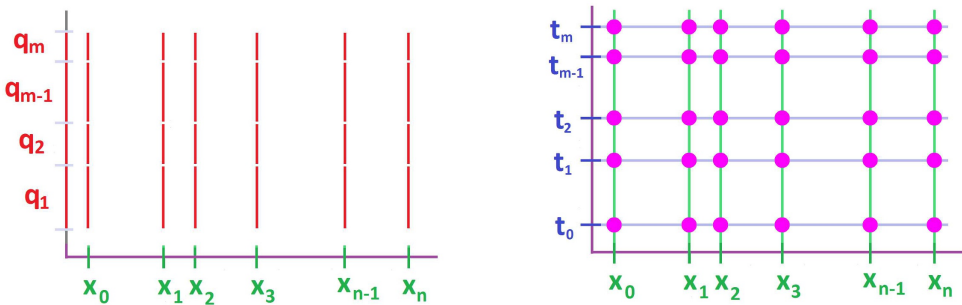
When m is constant, we have:

$$\frac{\Delta_t u(t, x)}{\Delta t} = m \frac{\Delta_x \frac{\Delta_x u(t, x)}{\Delta x}}{\Delta x}.$$

Or, in the simplified form:

$$\frac{\Delta_t u}{\Delta t} = m \frac{\Delta_x^2 u}{\Delta x^2}$$

We should keep in mind that the left-hand side is defined at the vertical edges and the right-hand side at the primary nodes of the decomposition of the rectangle.



Exercise 6.4.7

Prove that if the temperatures change linearly along the rod ($\Delta x = \Delta x = 1$), then they don't change with time (steady state).

6.5. The heat PDE

We now consider the *continuous case* of heat transfer.

Suppose the temperature function u is defined for *all* x and t within some open subset U of the plane and it is sampled at the nodes of a cell decomposition of the infinite rectangle $[a, b] \times [0, \infty)$ contained in that subset. We refine this cell decomposition of the rectangle and take the limit of the discrete heat equation with constant permeability k ,

$$\frac{\Delta u}{\Delta t}(q_i, x_k) = K \frac{\Delta^2 u}{\Delta x^2}(t_i, x_k),$$

as $\Delta x \rightarrow 0$ and $\Delta t \rightarrow 0$. We have the *heat equation*:

$$\frac{\partial u}{\partial t} = K \frac{\partial^2 u}{\partial x^2}$$

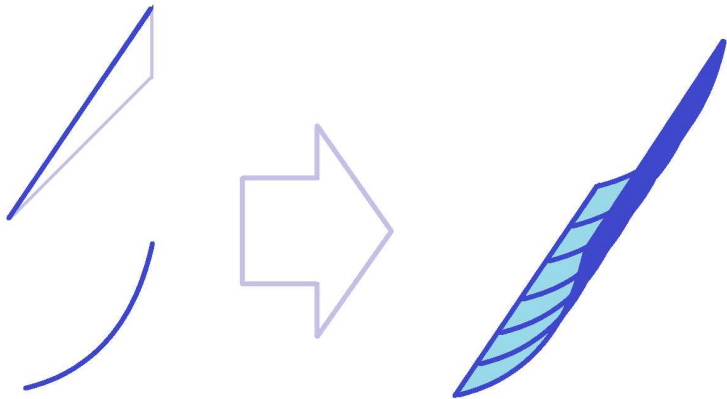
A solution u of this partial differential equation (PDE) is a function:

- defined and continuous on the rectangle,
- differentiable with respect to t inside of it, and
- twice differentiable with respect to x inside the rectangle, such that
- the equation is satisfied for each pair (x, t) such that x is in (a, b) and t is in $(0, \infty)$.

The meaning of the PDE is as follows:

- A positive concavity with respect to x goes together with a positive slope with respect to t .
- A negative concavity with respect to x goes together with a negative slope with respect to t .

In other words, we have the two patterns for u one of which is shown below:



Because this is a function of two variables, a solution is not made specific by a single-number initial condition as we know. We impose a combination of:

- the *initial condition*, providing the values of u in the beginning:

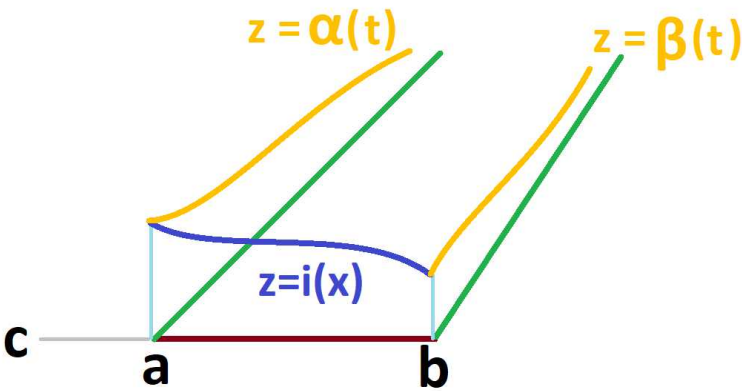
$$u(0, x) = i(x), \quad \text{when } a \leq x \leq b,$$

where the function i is given, and

- the *boundary condition*, providing the values of u at the ends of the rod:

$$u(t, a) = \alpha(t) \quad \text{and} \quad u(t, b) = \beta(t), \quad \text{when } t \geq 0,$$

where the functions α and β are given.



Exercise 6.5.1

Under the above restrictions what can you say about $i(x)$?

Exercise 6.5.2

Describe the model setups given in the previous sections with your choices of i , α , and β .

Example 6.5.3: steady state

A steady state solution is one that doesn't change with time:

$$u(t, x) = s(x) ,$$

for some function s . In other words,

$$u_t = 0 .$$

It follows from the PDE that

$$u_{xx} = 0 .$$

Therefore, $u(t, x) = s(x)$ is linear on x . If the initial state i is linear, the system will remain in it.

Example 6.5.4: maintained temperature

Suppose the temperatures at the ends are constant and equal:

$$\alpha = \beta = 0 .$$

That's the room temperature! As modeling has shown, the temperatures in the rod will approach this number:

$$u(t, x) \rightarrow 0 \text{ as } t \rightarrow \infty .$$

From the original Newton's Law, we know that this convergence is expected to be exponential. Therefore, we should try for $K = 1$:

$$u(t, x) = i(x)e^{-t} .$$

We substitute this into the PDE:

$$-i(x)e^{-t} = i''(x)e^{-t} .$$

Then,

$$-i(x) = i''(x) .$$

That's a familiar second order ODE from Chapter 1! The solution is:

$$i(x) = A \sin x + B \cos x .$$

So, this is possible!

Now the general case.

The idea is:

► What if u is the product of the initial state i by an exponential decay function of t ?

Let's test this idea by assuming that it's true:

$$u(t, x) = i(x)g(t) .$$

Substituting u into the PDE gives us the following:

$$ig' = Ki''g .$$

Let's rearrange the terms and "separate the variables":

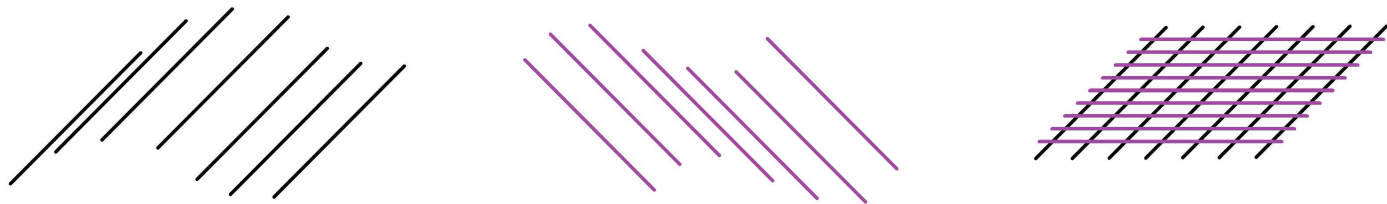
$$\frac{g'}{Kg} = \frac{i''}{i} .$$

Let's examine the equation:

$$\frac{g'(t)}{Kg(t)} = \frac{i''(x)}{i(x)}.$$

The left-hand side depends only on t is, therefore, constant with respect to x . Conversely, right-hand side depends only on x is, therefore, constant with respect to t .

What kind of function is this? One can imagine a roof to be built from vertical rods. Various shapes are possible when the rods are pointed in the same direction (left and middle):



The only way to have a cross pattern is a flat roof (right).

Therefore, this quantity must be constant with respect to both t and x ! Both sides are equal to some constant number, say $-\lambda$. We have:

$$g'(t) = -\lambda K g(t) \text{ and } i''(x) = -\lambda i(x).$$

We can solve these familiar ODEs.

The first one is the population ODE (Chapter 1). Its solution is:

$$g(t) = Ce^{-\lambda Kt}.$$

Since we expect exponential decay, we will consider only the case:

$$\lambda > 0.$$

With this in mind, we solve the second ODE. The solution is:

$$i(x) = A \sin(\sqrt{\lambda} x) + B \cos(\sqrt{\lambda} x),$$

for some constant A, B .

We conclude the following.

Theorem 6.5.5: Solutions of Heat PDE

The function

$$u(t, x) = e^{-\lambda Kt} \left(A \sin(\sqrt{\lambda} x) + B e^{-\lambda Kt} \cos(\sqrt{\lambda} x) \right)$$

is a solution to the heat equation.

For further analysis, we make two simplifying assumptions. First, we assume that the time starts at 0:

$$a = 0,$$

Second, we assume the *zero boundary condition*:

$$\alpha(t) = \beta(t) = 0.$$

Then the zero function is always a solution.

The boundary conditions produce:

$$u(0,t)=i(0)g(t)=0\implies i(0)=0\implies A\sin(\sqrt{\lambda}0)+B\cos(\sqrt{\lambda}0)=0\implies B=0,$$

and

$$u(t,b)=i(b)g(t)=0\implies i(b)=0\implies A\sin(\sqrt{\lambda}b)+0\cos(\sqrt{\lambda}b)=0\implies \sqrt{\lambda}b=\pi n,$$

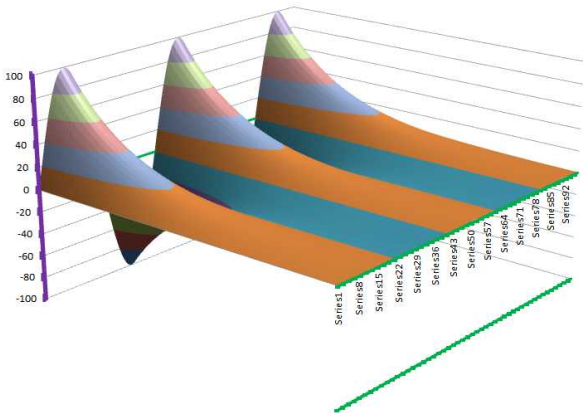
for some integer n . We have proven the following:

Theorem 6.5.6: Solutions of Heat PDE With Zero Boundary Condition

If
$$\sqrt{\lambda}=\pi n/b$$
 for some integer n , then there is a solution to the heat equation with the zero boundary condition:

$$u_n(t,x)=Be^{-\lambda Kt}\sin(\sqrt{\lambda}x)$$

For each t , this is a horizontally and vertically stretched/shrunk sinusoid. As t grows, the magnitude of its amplitude diminishes but the period remains the same. The function above is plotted below:



An important observation is the following:

Corollary 6.5.7: Sum of Solutions of the Heat PDE

The sum of solutions of the heat equation is also a solution of the heat equation; therefore,

$$u(t,x)=\sum_{n=1}^NB_ne^{-\lambda_nKt}\sin(\sqrt{\lambda_n}x)$$

where

$$\sqrt{\lambda_n}=\pi n/b,\;n=1,2,\dots,N,$$

is a solution of the heat equation with the zero boundary condition.

Exercise 6.5.8

Prove the corollary.

We recognize the partial sums of the Fourier series. When convergent, its sum is also a solution.

6.6. Cells and forms in higher dimensions

We used cell decompositions of intervals, as well as those of the whole real line, in order to study *incremental change*. This time, we need cell decompositions of the n -dimensional Euclidean space. The building blocks will come from cell decompositions of the axes.

For dimension 2, these are *rectangles*. An interval in the x -axis:

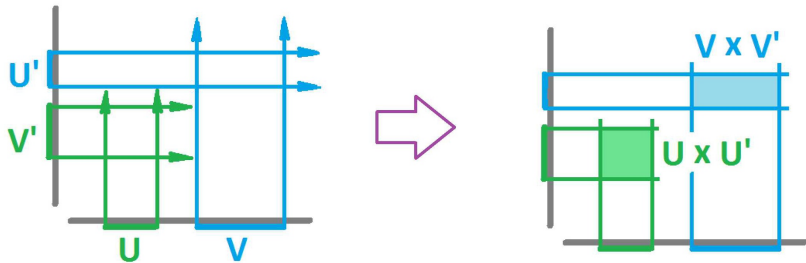
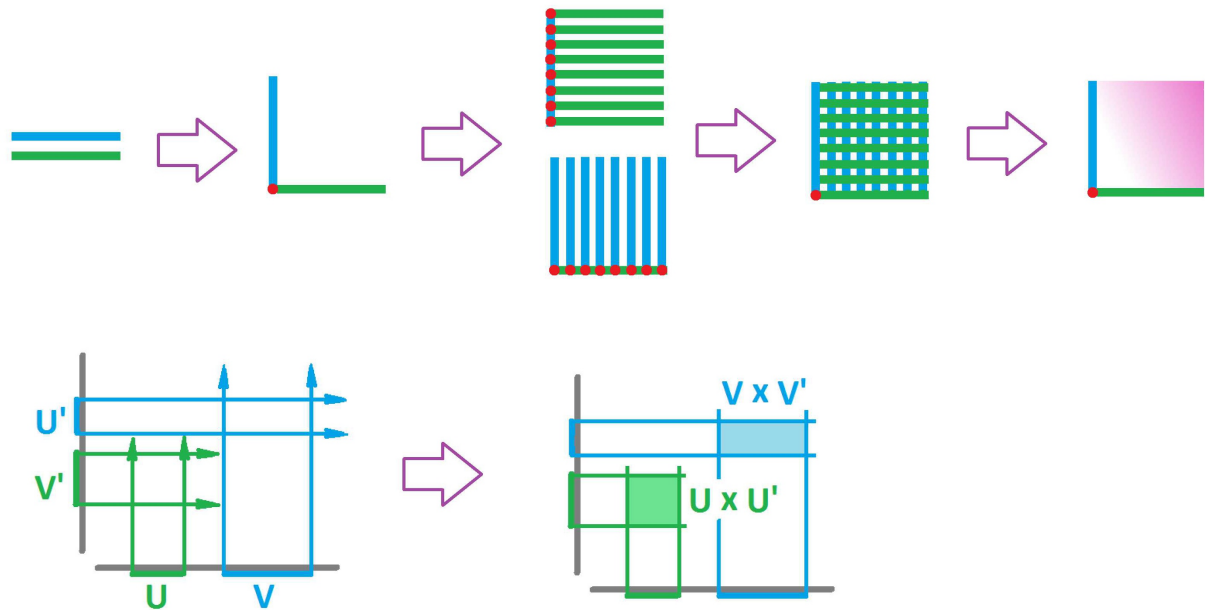
$$[a,b] = \{x : a \leq x \leq b\},$$

and an interval in the y -axis:

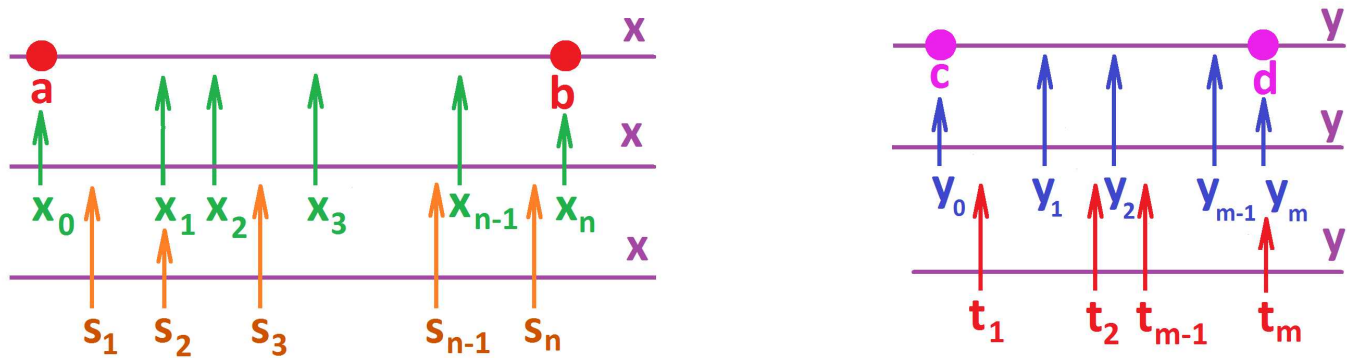
$$[c,d] = \{y : c \leq y \leq d\},$$

make a rectangle in the xy -plane:

$$[a,b] \times [c,d] = \{(x,y) : a \leq x \leq b, c \leq y \leq d\}.$$



A *cell decomposition* of the rectangle $[a,b] \times [c,d]$ is made of smaller rectangles constructed in the same way as above. Suppose we have cell decompositions of the intervals $[a,b]$ in the x -axis and $[c,d]$ in the y -axis:



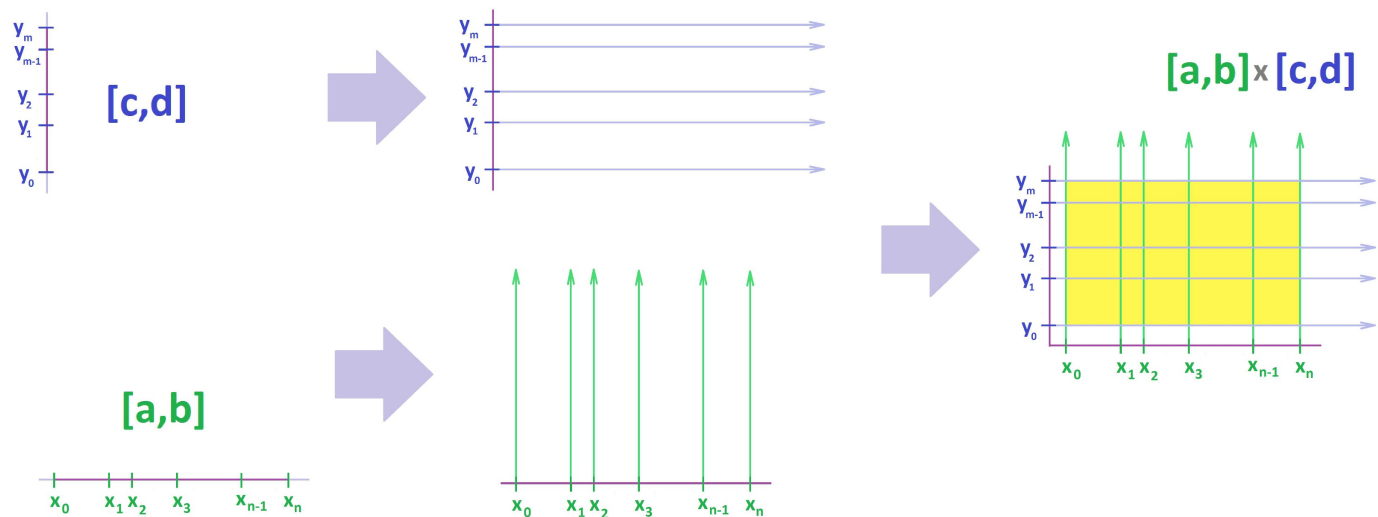
We start with a cell decomposition of an interval $[a,b]$ in the x -axis into n intervals:

$$[x_0,x_1], [x_1,x_2], \dots, [x_{n-1},x_n],$$

with $x_0 = a, x_n = b$. Then we do the same for y . We decompose an interval $[c,d]$ in the y -axis into m intervals:

$$[y_0,y_1], [y_1,y_2], \dots, [y_{m-1},y_m],$$

with $y_0 = c, y_m = d$.

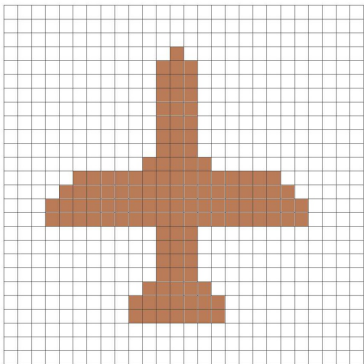


The lines $y = y_j$ and $x = x_i$ create a cell decomposition of the rectangle $[a, b] \times [c, d]$ into smaller rectangles $[x_i, x_{i+1}] \times [y_j, y_{j+1}]$. The points of intersection of these lines,

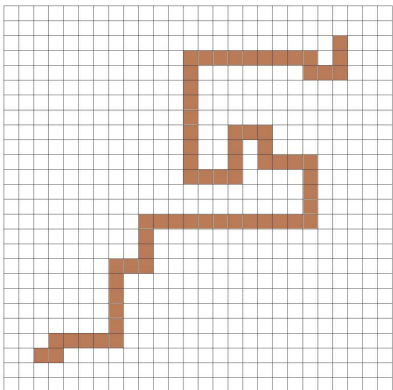
$$X_{ij} = (x_i, y_j), \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, m,$$

will be called the *nodes* of the cell decomposition. So, there are nodes and there are rectangles (tiles); is that it?

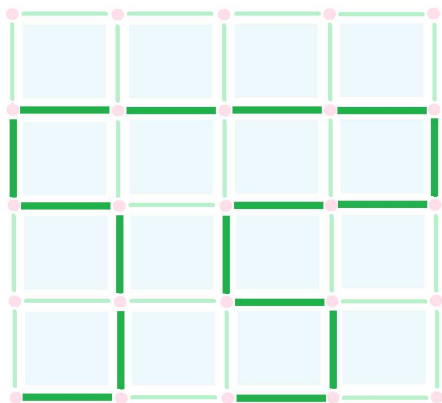
This is how an object can be represented with tiles, or pixels:



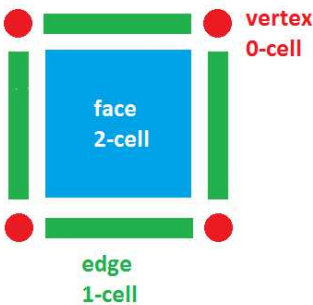
Now, are *curves* also made of tiles? Such a curve would look like this:



If we look closer, however, this “curve” isn’t a curve in the usual sense; it’s thick! The correct answer is: *curves are made of edges* of the grid:



We have discovered that we need to include, in addition to the squares, the “thinner” cells as additional building blocks. The complete decomposition of the pixel is shown below; the edges and vertices are shared with adjacent pixels:



Example 6.6.1: dimension 1

We start with dimension $n = 1$:

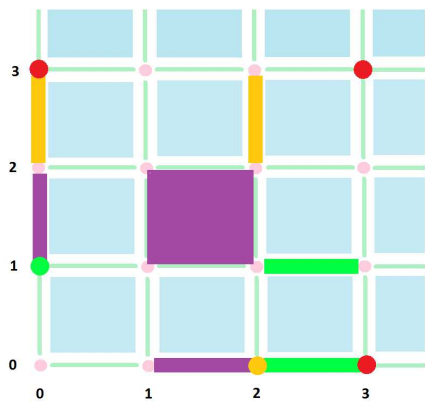


- In this simplest of cell decompositions, the cells are:
- A node, or a 0-cell, is $\{k\}$ with $k = \dots, -2, -1, 0, 1, 2, 3, \dots$
 - An edge, or a 1-cell, is $[k, k + 1]$ with $k = \dots, -2, -1, 0, 1, 2, 3, \dots$
 - And, 1-cells are attached to each other along these 0-cells.

Example 6.6.2: dimension 2

For the dimension $n = 2$ grid, we define cells for all integers k, m as products:

- A node, or a 0-cell, is $\{k\} \times \{m\}$.
 - An edge, or a 1-cell, is $\{k\} \times [m, m + 1]$ or $[k, k + 1] \times \{m\}$.
 - A square, or a 2-cell, is $[k, k + 1] \times [m, m + 1]$.
- We also have:
- The 2-cells are attached to each other along these 1-cells.
 - And, still, the 1-cells are attached to each other along the 0-cells.



- Cells shown above are:
- 0-cell $\{3\} \times \{3\}$,
 - 1-cells $[2, 3] \times \{1\}$ and $\{2\} \times [2, 3]$,
 - 2-cell $[1, 2] \times [1, 2]$.

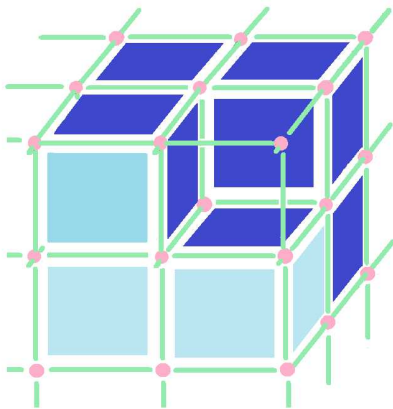
Similarly for dimension 3, we have *boxes*. Intervals in the x -, y -, and z -axes:

$$[a, b] = \{x : a \leq x \leq b\}, \quad [c, d] = \{y : c \leq y \leq d\}, \quad [p, q] = \{z : p \leq z \leq q\},$$

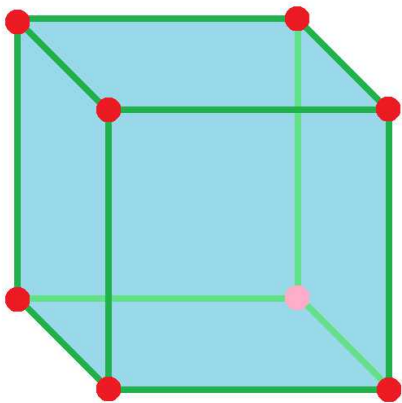
make a box in the xyz -space:

$$[a, b] \times [c, d] \times [p, q] = \{(x, y) : a \leq x \leq b, \; c \leq y \leq d, \; p \leq z \leq q\}.$$

In dimension 3, *surfaces are made of faces* of our boxes; i.e., these are tiles:



The cell decomposition of the box follows and here, once again, the faces, edges, and vertices are shared:

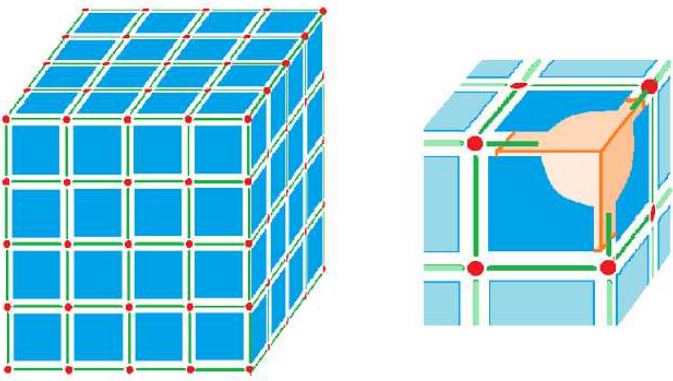


Example 6.6.3: dimension 3

For all integers i, m, k , we have:

- A node, or a 0-cell, is $\{i\} \times \{m\} \times \{k\}$.

- An edge, or a 1-cell, is $\{i\} \times [m, m + 1] \times \{k\}$, etc.
- A square, or a 2-cell, is $[i, i + 1] \times [m, m + 1] \times \{k\}$, etc.
- A cube, or a 3-cell, is $[i, i + 1] \times [m, m + 1] \times [k, k + 1]$.



Thus, our approach to decomposition of space, in any dimension, boils down to the following:

► *The n -dimensional space is composed of cells in such a way that k -cells are attached to each other along $(k - 1)$ -cells, $k = 1, 2, \dots, n$.*

The examples show how the n -dimensional Euclidean space is decomposed into 0-, 1-, ..., n -cells in such a way that

- n -cells are attached to each other along $(n - 1)$ -cells.
- $(n - 1)$ -cells are attached to each other along $(n - 2)$ -cells.
- ...
- 1-cells are attached to each other along 0-cells.

What are those cells exactly?

Definition 6.6.4: cell

In the n -dimensional space, \mathbf{R}^n , a *cell* is the subset given by the product with n components:

$$P = I_1 \times \dots \times I_n,$$

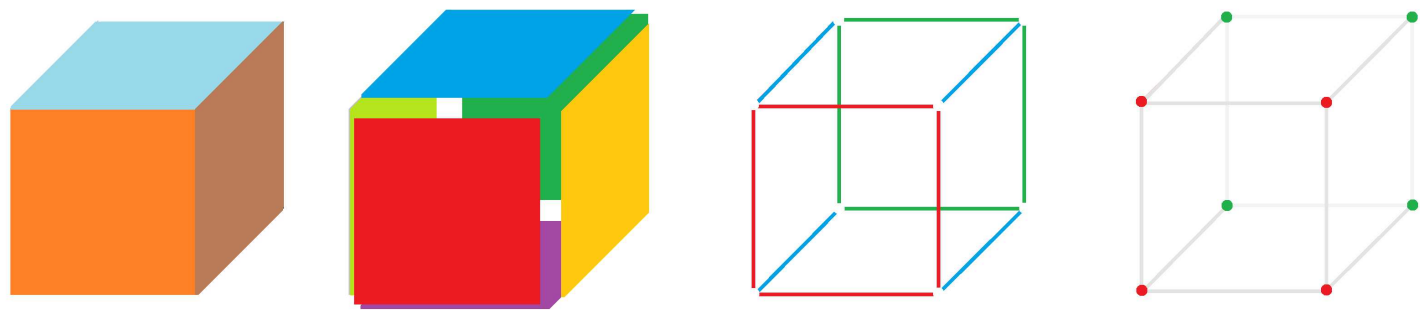
with its k th component is either

- a closed interval $I_k = [x_k, x_{k+1}]$, or
- a point $I_k = \{x_k\}$.

The cell's *dimension* is equal to m , and it is also called an m -cell, when there are m edges and $m - n$ vertices on this list. Replacing one of the edges in the product with one of its end-points creates an $(n - 1)$ -cell called a *boundary cell* of P .

Definition 6.6.5: face

Replacing one of the edges in the product with one of its end-points creates an $(n - 1)$ -cell called a *face* of P . Replacing several edges with one of their end-points creates an k -cell, $k < n$, called a *boundary cell* of P .



Thus, cell decompositions of the axes – into nodes and edges – create a cell decomposition of the whole space – into cells of all dimensions.

Example 6.6.6: 3d decomposition

Below, a 3-cell is shown as a “room” along with all of the cells of dimensions 0, 1, 2:

Joints and beams

Four walls

Ceiling and floor

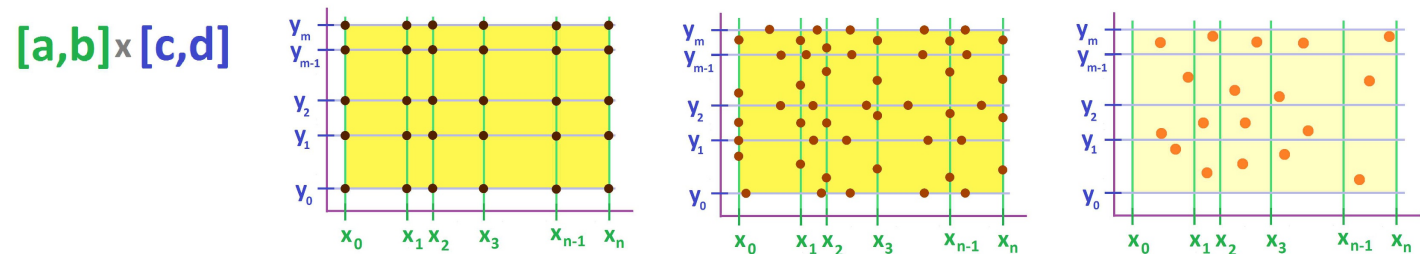
Room

They all come from the nodes and edges on the axes:

- 0: Each of the joints of the “beams” is the product of three nodes.
- 1: Each of the “beams” is the product of two nodes and an edge.
- 2: Each of the “walls”, as well as the “floor” and the “ceiling”, is the product of two edges and a node.
- 3: The “room” is the product of three edges.

The 2-cells here are the faces of the 3-cell, the 1-cells are the faces of the 1-cells, etc.

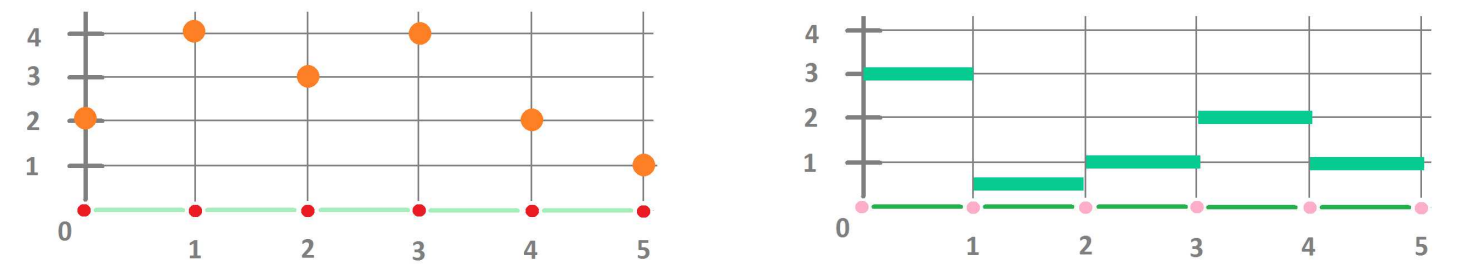
In Volumes 2 and 3, we assigned numbers to *points* within cells in the 1-dimensional case to represent such things as location – nodes or 0-cells – and velocity – secondary nodes or 1-cells. We will continue to do so. In fact, we will study functions defined at points located at the cells of a particular dimension m in a cell decomposition. Below we see $m = 0, 1, 2$, respectively:



Firstly, these points – secondary, tertiary, etc. nodes – may be specified as a result of *sampling* a function defined on the whole region. Note that, in that case, one node may be shared by several adjacent cells.

Secondly, these points are used for mere *bookkeeping*. We then can choose them to be the end-points or corners or mid-points etc. In truth though, the quantities are assigned to the cells *themselves*. In other words, each cell is an input of these functions, as explained below.

Recall how we defined discrete forms for dimension 1: within each of the pieces of a cell decomposition of the line this function is unchanged; i.e., it’s a single number. This is how we plot the graphs of 0- and 1-forms over \mathbf{R}^1 :



There are more types of cells in the higher dimensional spaces, but the idea remains:

Where forms live...

0-forms

1-forms

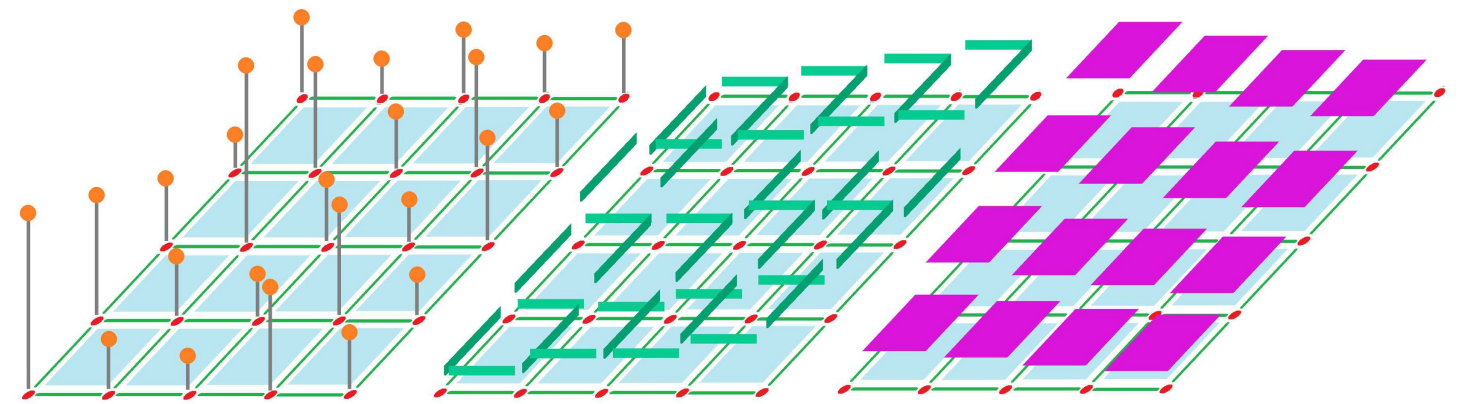
1-forms

2-forms

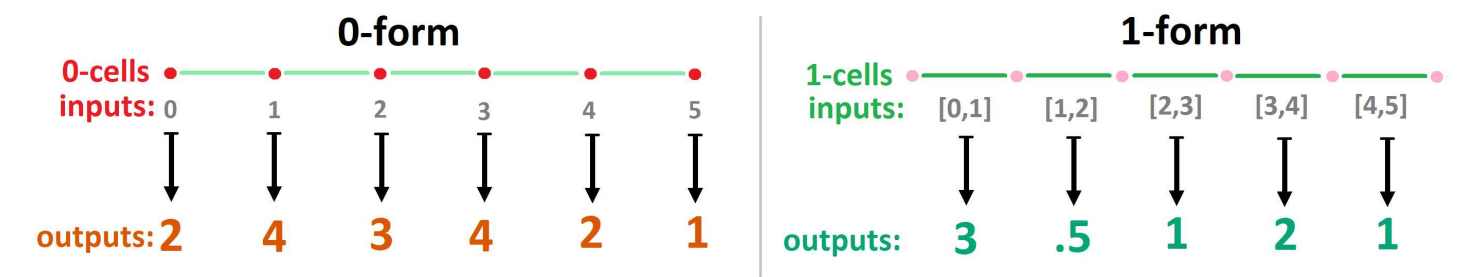
Definition 6.6.7: discrete form

A *discrete form of degree k over \mathbf{R}^n* , or simply a k -form, is a real-valued function defined on k -cells of \mathbf{R}^n .

And these are 0-, 1-, and 2-forms over \mathbf{R}^2 :



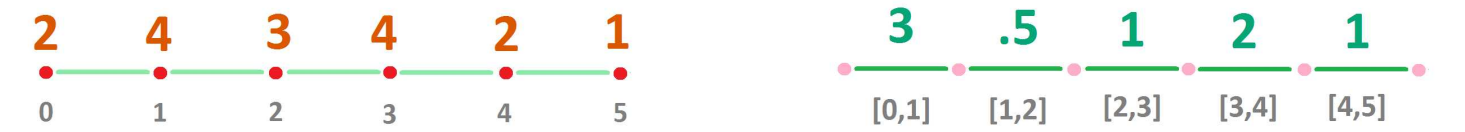
To emphasize the nature of a form as a function, we can use arrows (\mathbf{R}^1):



Here we have two forms:

- a 0-form with $0 \mapsto 2, 1 \mapsto 4, 2 \mapsto 3, \dots$; and
- a 1-form with $[0, 1] \mapsto 3, [1, 2] \mapsto .5, [2, 3] \mapsto 1, \dots$

A more compact way to visualize is this:



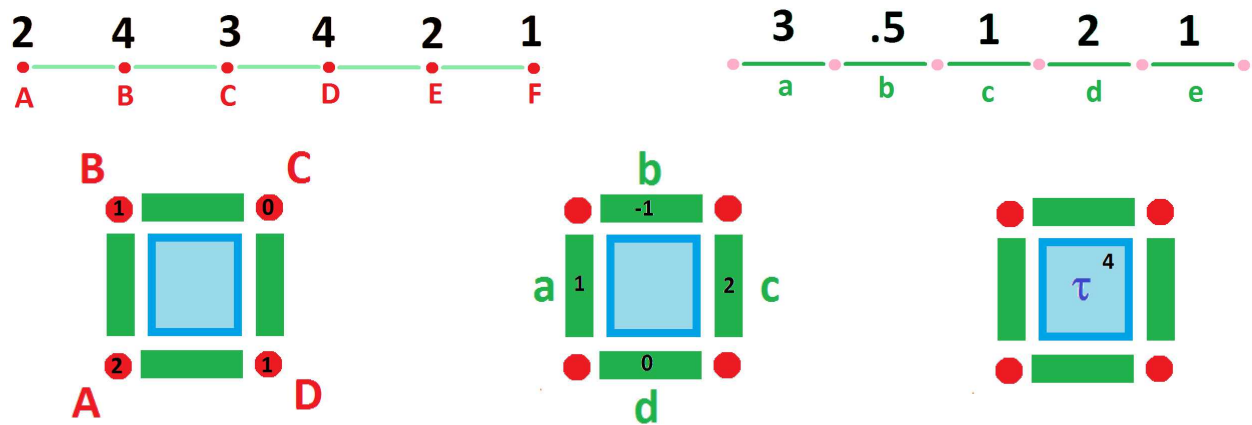
Here we have two forms:

- a 0-form q with $q(0) = 2, q(1) = 4, q(2) = 3, \dots$; and
- a 1-form s with $s([0, 1]) = 3, s([1, 2]) = .5, s([2, 3]) = 1, \dots$

We can also use letters to label the cells, just as before. Each cell is then assigned *two* symbols:

- one is its name (a latter) and
- the other is the value of the form at that location (a number).

This idea is illustrated for forms over \mathbf{R}^1 and \mathbf{R}^2 respectively:



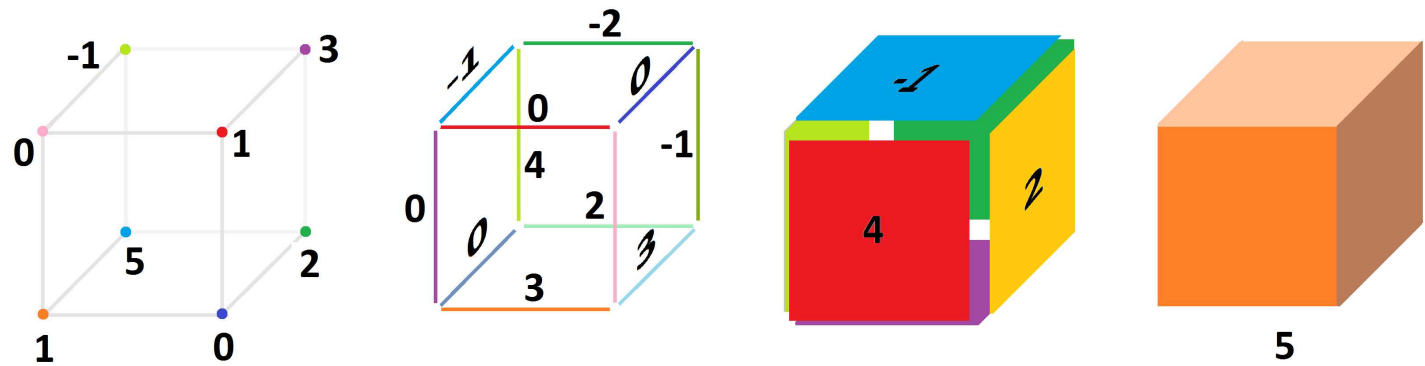
We have a 0-form q and a 1-form s in the former example:

- $q(A) = 2, q(B) = 4, q(C) = 3, \dots$
- $s(AB) = 3, s(BC) = .5, s(CD) = 1, \dots$

We also have a 2-form ϕ in the latter example:

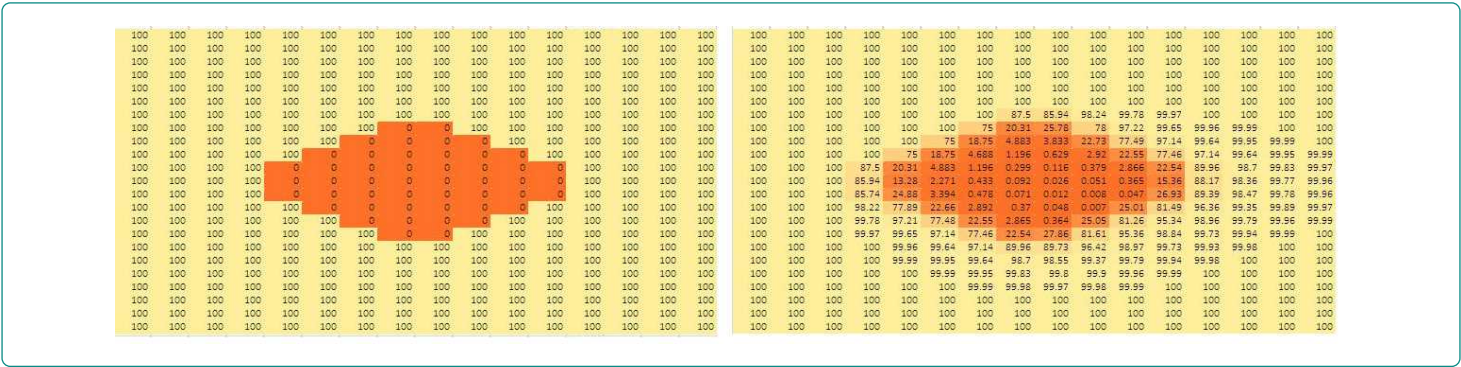
- $q(A) = 2, q(B) = 1, q(C) = 0, q(D) = 1$
- $s(a) = 1, s(b) = -1, s(c) = 2, s(d) = 0$
- $\phi(\tau) = 4$

We can simply label the cells with numbers, as follows (in \mathbf{R}^3):



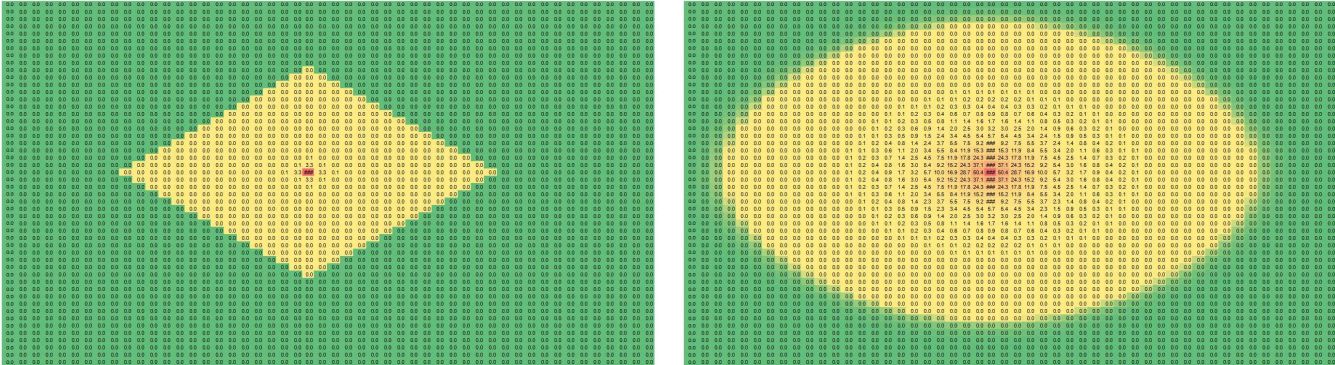
These forms may represent the following characteristics of a flow of a liquid:

- A 0-form: the pressure of the liquid at the joints of a system of pipes.
- A 1-form: the flow rate of the liquid along the pipe.



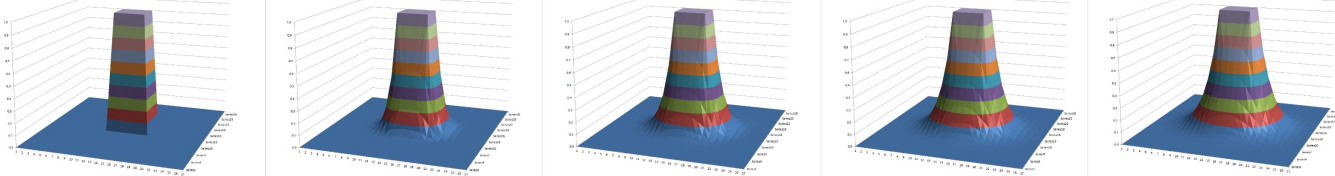
Example 6.7.2: a source of heat

On a larger scale, the simulation produces a realistic circular pattern even though it starts straight. Below we have a single-point but permanent source of heat shown after 3 and then after 3000 iterations:



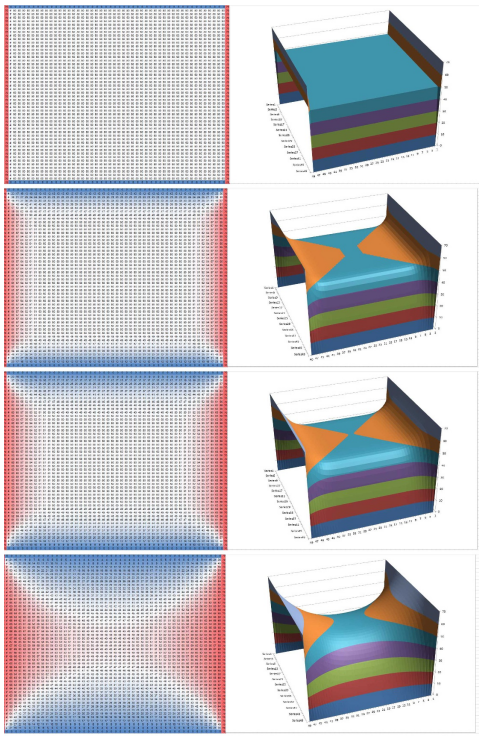
Example 6.7.3: heater

For each t , our function $u(t, \cdot, \cdot)$ is a function of two variables and can be visualized by its graph. For example, a square heater in the middle of a room will produce the following dynamics of distribution of the temperature over the floor:



Example 6.7.4: walls

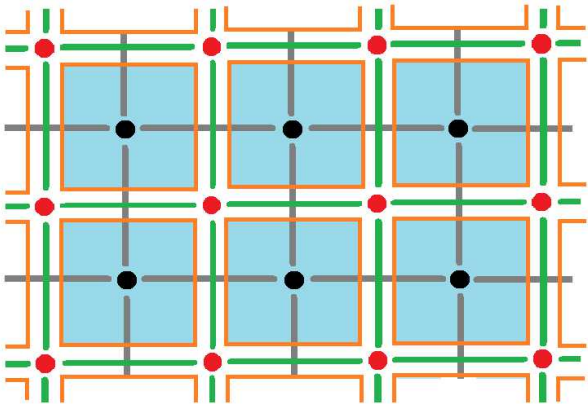
Two opposite walls of a room are warm and the other two are cold. The temperature develops into a saddle-like pattern:



Exercise 6.7.5

Implement the above examples.

Thus, we have *four* walls of the room or, dually, four pipes leaving the room and that is how the heat (or material) is exchanged with the neighbors. This correspondence is illustrated below.



The latter interpretation is preferable because our temperature distribution function w is then a 0-form in a space of any dimension.

In contrast to the above simulations, for the general case, we will take into account the permeability of the walls/pipes.

Recall that a cell decomposition of a *box* B in the txy -space comes from cell decompositions of its three edges as described in [Chapter 4HD-5](#):

$$\begin{array}{cccccccc} t_0 & q_1 & t_1 & q_2 & t_2 & q_3 & \dots \\ x_0 & s_1 & x_1 & s_2 & x_2 & s_3 & \dots \\ y_0 & p_1 & y_1 & p_2 & y_2 & p_3 & \dots \end{array}$$

We make a simplifying assumption that all containers have equal sizes. Then, for container located at

(x_i, y_j) , this is the *total inflow*:

•

$-K(t_k, s_{i-1}, y_j)\Delta_x u(t_k, s_{i-1}, y_j)$

•

$-K(t_k, s_i, y_{j-1})\Delta_y u(t_k, s_i, p_{j-1})$

$+K(t_k, s_i, y_j)\Delta_y u(t_k, s_i, p_j)$

•

$+K(t_k, s_i, y_j)\Delta_x u(t_k, s_i, y_j)$

•

The four terms are the inflows across each of the four walls of the container and they are arranged accordingly. The proportion of the heat exchanged across each wall is given by the function $K \geq 0$. It incorporates the permeability of the walls, their sizes, the length of the time interval, and so on.

The *difference* equation for the heat, or temperature, $u = u(t, x, y)$ is the following:

$$\Delta_t u(q_k, x_i, y_j) = \text{total inflow} \text{ ,}$$

and the recursive formula to be implemented is simply:

$$u(t_k, x_i, y_j) = u(t_{k-1}, x_i, y_j) + \text{total inflow} \text{ .}$$

The spreadsheet consists of several sheets computed consecutively:

- the permeability for each wall,
- the initial temperature for each container,
- the buffer (copied current values),
- the difference and the flow of temperature for each wall,
- the total flow for each container, and finally
- the current values of the temperature.

The last formula is:

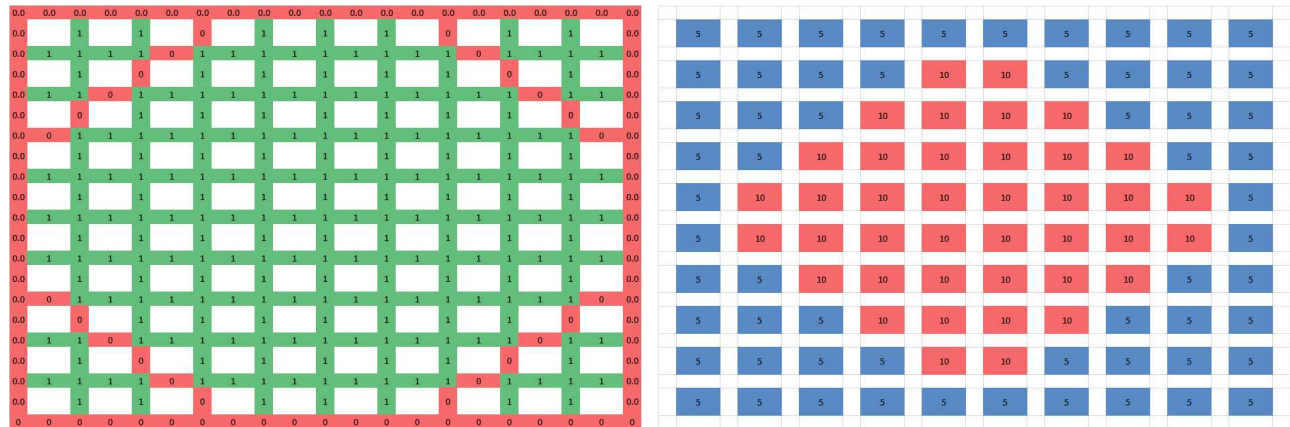
the current value = the buffer value + the total inflow .

	=IF(R1C1=0,initialIRC,bufferIRC+R1C5*fluxIRC)																											
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26		
1	2	Δx	0.10	k+	-0.1																							
2	3	Δy	0.10	y+	0.0	0.1	0.1	0.2	**	0.3	**	0.4	**	0.5	**	0.6	**	0.7	**	0.8	**	0.9	**	1.0	1.0	1.1		
3	4				0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
4	5		0.00																									
5	6	total	0		5.00	5.00	5.00	5.07	5.07	5.07	5.07	5.07	5.07	5.07	5.07	5.07	5.07	5.07	5.07	5.07	5.07	5.07	5.07	5.07	5.07	5.07		
6	7	505.00	0.10		5.00	5.00	5.06	5.07	5.07	5.07	5.07	5.07	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08		
7	8		0.15	0	5.00	5.00	5.06	5.07	5.07	5.07	5.07	5.07	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08		
8	9		0.20		5.00	5.05	5.06	5.07	5.07	5.07	5.07	5.07	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08		
9	10		0.25	0	5.00	5.05	5.06	5.07	5.07	5.07	5.07	5.07	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08		
10	11		0.30		5.00	5.05	5.06	5.07	5.07	5.07	5.07	5.07	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08		
11	12		0.35	0	5.04	5.05	5.06	5.07	5.07	5.07	5.07	5.07	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.07		
12	13		0.40		5.04	5.05	5.06	5.07	5.07	5.07	5.07	5.07	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.07		
13	14		0.45	0	5.04	5.05	5.06	5.07	5.07	5.07	5.07	5.07	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.07		
14	15		0.50		5.04	5.05	5.06	5.07	5.07	5.07	5.07	5.07	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.07		
15	16		0.55	0	5.04	5.05	5.06	5.07	5.07	5.07	5.07	5.07	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.07		
16	17		0.60		5.04	5.05	5.06	5.07	5.07	5.07	5.07	5.07	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.08	5.07		
17	18		0.65	0	5.04	5.04	5.05	5.06	5.07	5.07	5.07	5.07	5.07	5.07	5.07	5.07	5.07	5.07	5.07	5.07	5.07	5.07	5.07	5.07	5.07	5.07		
18	19		0.70		5.00	5.04	5.05	5.05	5.05	5.05	5.05	5.05	5.06	5.06	5.06	5.06	5.06	5.06	5.06	5.06	5.06	5.06	5.06	5.06	5.06	5.00		
19	20		0.75	0	5.00	5.04	5.05	5.05	5.05	5.05	5.05	5.05	5.06	5.06	5.06	5.06	5.06	5.06	5.06	5.06	5.06	5.06	5.06	5.06	5.06	5.00		
20	21		0.80		5.00	5.00	5.04	5.04	5.04	5.04	5.04	5.04	5.05	5.05	5.05	5.05	5.05	5.05	5.05	5.05	5.05	5.05	5.05	5.05	5.05	5.00		
21	22		0.85	0	5.00	5.00	5.04	5.04	5.04	5.04	5.04	5.04	5.05	5.05	5.05	5.05	5.05	5.05	5.05	5.05	5.05	5.05	5.05	5.05	5.05	5.00		
22	23		0.90		5.00	5.00	5.00	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.00		
23	24		0.95	0	5.00	5.00	5.00	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.00		
24	25		1.00		5.00	5.00	5.00	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.00		
25	26				5.00	5.00	5.00	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.04	5.00		
		perm	initial	buffer	flow	flux	values																					

As you can see, the square cells represent the containers and the narrow ones the walls.

Example 6.7.6: coffee

Suppose coffee is poured into an insulated cup. The temperature of the areas adjacent to the sides of the cup quickly cools down. At this point we start our simulation. These are the first two sheets: the permeability (zero around the edge of the cup) and the initial temperature (hot inside, cold outside):



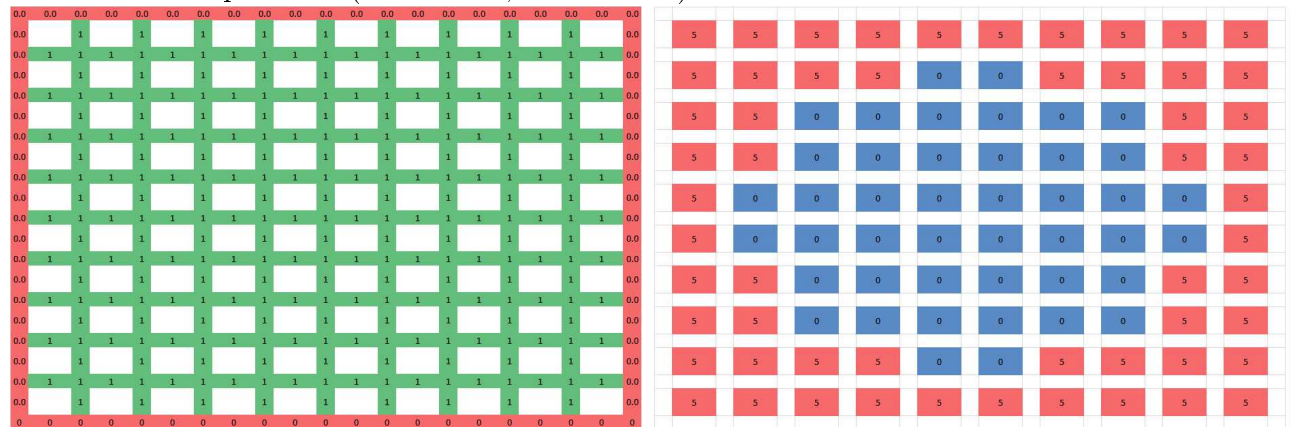
Heat transfer continues within the body of the coffee with virtually no transfer through the walls of the cup. These are the results after 50 iterations:



One can see that the total amount of heat is preserved.

Example 6.7.7: soda

Suppose a can of soda is taken out of the refrigerator. The areas adjacent to the sides of the can start to exchange heat with the outside. However, the effect on the outside temperature is negligible. At this point we start our simulation. These are the first two sheets: the permeability (high throughout) and the initial temperature (cold inside, hot outside):



The temperature outside the can remains unchanged! These are the results after 50 iterations:



Exercise 6.7.8

What if the permeability of a wall is proportional to the average temperature of the two adjacent containers?

When K is constant as another simplifying assumption, the right-hand side of our equation becomes recognizable:

$$\begin{aligned} \Delta_t u(q_k, x_i, y_j) &= K \left(\begin{array}{ccc} \bullet & -\Delta_y u(t_k, s_i, p_{j-1}) & \bullet \\ -\Delta_x u(t_k, s_{i-1}, y_j) & & +\Delta_x u(t_k, s_i, y_j) \\ \bullet & +\Delta_y u(t_k, s_i, p_j) & \bullet \end{array} \right) \\ &= K \left([\Delta_x u(t_k, s_i, y_j) - \Delta_x u(t_k, s_{i-1}, y_j)] + [\Delta_y u(t_k, s_i, p_j) - \Delta_y u(t_k, s_i, p_{j-1})] \right) \\ &= K (\Delta_x \Delta_x u(t_k, x_i, y_j) + \Delta_y \Delta_y u(t_k, x_i, y_j)) \\ &= K (\Delta_x^2 u(t_k, x_i, y_j) + \Delta_y^2 u(t_k, x_i, y_j)). \end{aligned}$$

These are the *second differences* introduced in [Chapter 4HD-3](#).

As the transfer is proportional to the length of the time interval, our *partial* difference equation becomes the following:

$$\Delta_t u = K (\Delta_x^2 u + \Delta_y^2 u) \Delta.$$

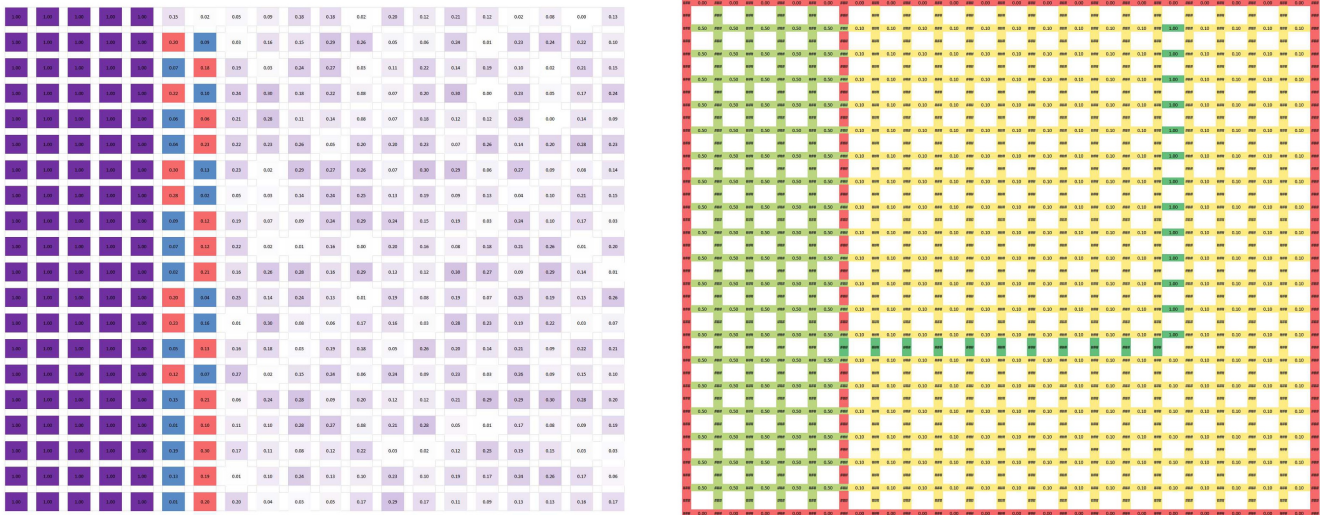
Exercise 6.7.9

Derive a version of this equation for a variable K .

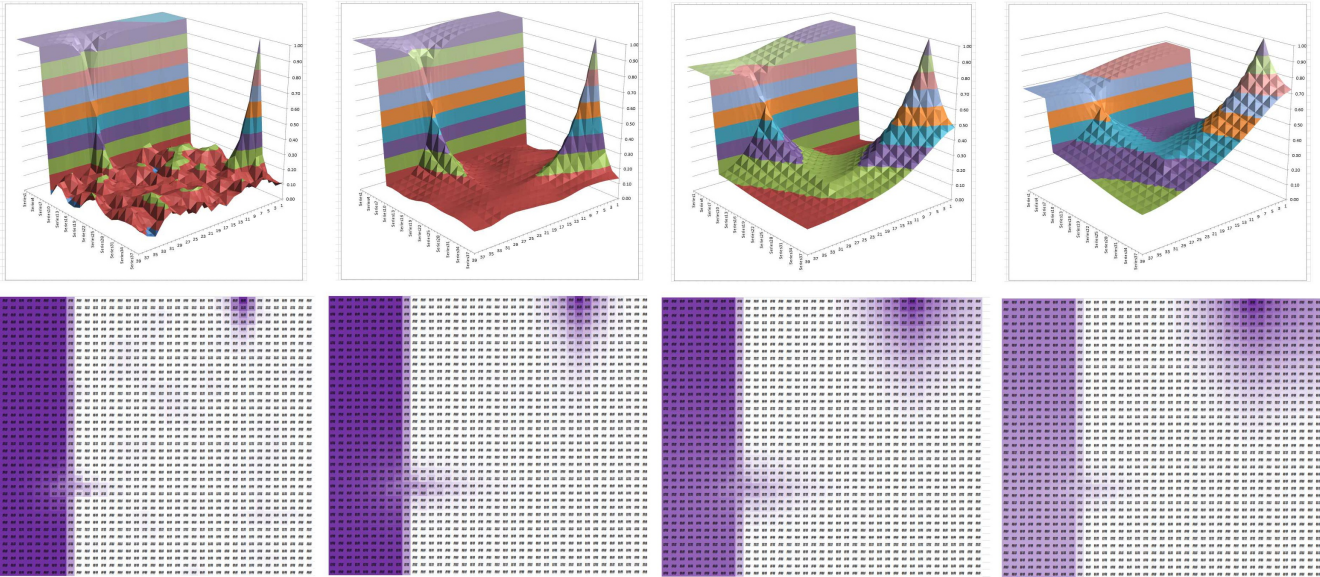
Example 6.7.10: migration

The discrete heat equation can also be used to model moving populations. Indeed, it is conceivable that people migrate from their current location to an adjacent area with a lower population density.

Below we imagine two areas in a country: one is densely and uniformly populated ($u(0, x, y) = 1$) and the other thinly and randomly ($u(0, x, y)$ close to 0). The two parts are separated by a river ($k = 0$). Now a bridge is built over the river as well a road ($k = 1$) from the bridge to a town on a city on the other end of the country. The initial distribution of the population is shown on the left and the permeability is shown on the right:



At this point, people start to cross the river and spread into the thinly populated area – especially along the road. We also imagine that the city is a constant source of new settlers arriving from the outside, $u(t, x_0, y_0) = 1$.



Exercise 6.7.11

(a) Incorporate into the model the possibility of growing population with location-dependent rates.

(b) Incorporate into the model the possibility of sustainability limits (location-dependent) on the population growth. Derive the PDE.

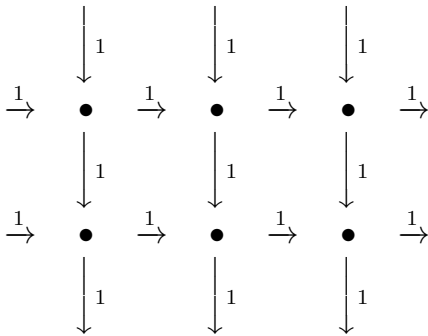
Example 6.7.12: diffusion in a flow

If we drop dye in a liquid, it will diffuse according to the heat equation. What if the liquid is in motion? Let’s see what happens if we let a drop of dye to be taken by a flow. As a simple example, we follow the rule that:

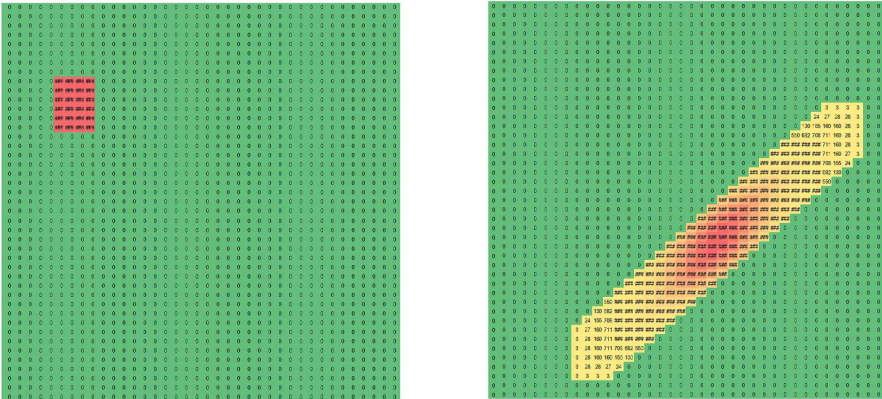
- The amount of dye in each cell is split and passed to the adjacent cells along the vectors of the vector field proportional to the values attached to the edges.

For example, we can choose 1’s on the horizontal edges and 0s on the vertical edges. Then the flow will be purely horizontal. If we reverse the values, it will be purely vertical. Now what if we choose

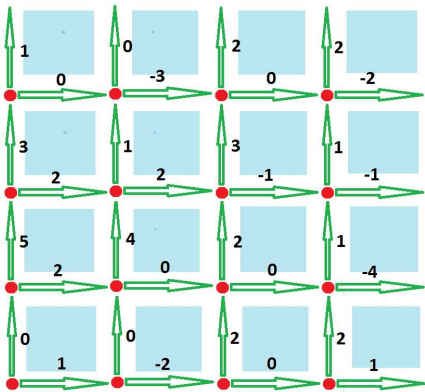
1's on both vertical and horizontal edges?



It is simple: the amount of dye in each cell is split in half and passed to the two adjacent cells along the vectors. The liquid is flowing down and right equally. We see some dispersal, but the predominantly diagonal direction of the spreading of the dye is also evident:



A more general example is this. Imagine that a liquid moves along the square grid which is thought of as a *system of pipes*:



In the picture above, the numbers represent “flows” through these “pipes”, with the direction determined by the direction of the x,y -axes. In particular,

- the “1” can be understood as: “1 cubic foot per second from left to right”.
 - the “2” can be understood as: “2 cubic feet per second upward”.
- However, for the sake of *conservation of matter*, these numbers have to be normalized. Then 1/3 of the amount goes right and 2/3 goes up. Of course, we this is a discrete 1-form.

Exercise 6.7.13

Create a spreadsheet and confirm the pattern. What happens if the flow takes horizontally twice as much material as vertically?

Following the development for dimension 1, we introduce the lengths of pipes and the dimensions of the cells into the models. The result is the second difference quotients in the right hand side:

$$\frac{\Delta u}{\Delta t} = \alpha \left(\frac{\Delta^2 u}{\Delta x^2} + \frac{\Delta^2 u}{\Delta y^2} \right) .$$

Their sum is called the Laplace operator.

6.8. The heat PDE for dimension 2

We now consider the *continuous case* of heat transfer.

Suppose the temperature function u is defined for *all* x , y , and t within some open subset U of the space and it is sampled at the nodes of a cell decomposition of the infinite box $[a, b] \times [c, d] \times [0, \infty)$ contained in that subset.

We refine this cell decomposition of the rectangle and take the limit of the discrete heat equation with constant permeability K ,

$$\frac{\Delta u}{\Delta t}(q_i, x_k, y_j) = K \left(\frac{\Delta^2 u}{\Delta x^2}(t_i, x_k, y_j) + \frac{\Delta^2 u}{\Delta y^2}(t_i, x_k, y_j) \right),$$

as

$$\Delta x \rightarrow 0, \Delta y \rightarrow 0, \text{ and } \Delta t \rightarrow 0.$$

We have the *heat equation*:

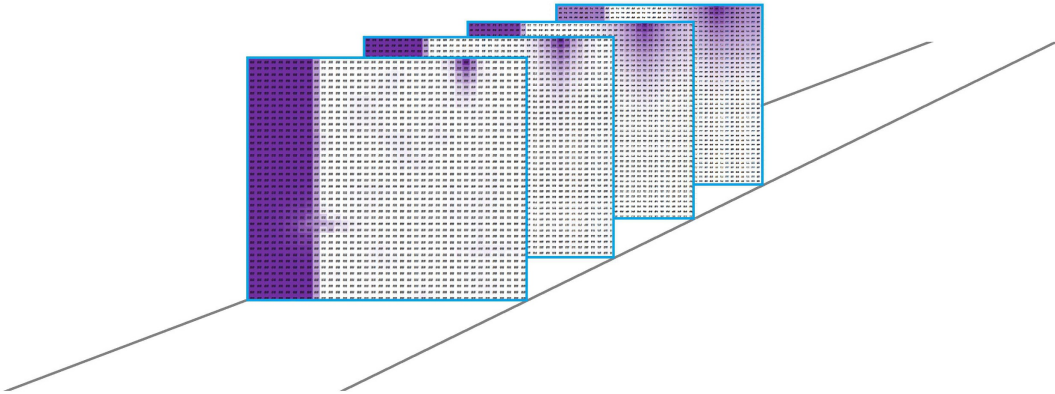
$$\frac{\partial u}{\partial t} = \alpha \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right)$$

The term in parentheses is called the *Laplace operator* of u .

A solution u of this partial differential equation (PDE) is a function

- defined and continuous on the box,
- differentiable with respect to t inside of it, and
- twice differentiable with respect to x and y inside the box, such that
- the equation is satisfied for each (t, x, y) .

A visualization of the solution is identical to the discrete case:



The meaning of the PDE is that:

- A positive concavity with respect to x or y goes together with a positive slope with respect to t .
- A negative concavity with respect to x or y goes together with a negative slope with respect to t .

We impose these conditions to make the solution specific:

- the *initial condition*, providing the values of u in the beginning:

$$u(0,x,y)=i(x,y), \quad \text{when } a \leq x \leq b, \; c \leq y \leq d$$

where the function i is given, and

- the *boundary condition*, providing the values of u at the edges of the rectangle for all $t \geq 0$:

$$u(t,a,y)=\alpha(t,y), \; u(t,b,y)=\beta(t,y) \quad \text{when } c \leq y \leq d,$$

and

$$u(t,x,c)=\gamma(t,y), \; u(t,x,d)=\delta(t,x) \quad \text{when } a \leq x \leq b,$$

where the functions α, β, γ , and δ are given.

Exercise 6.8.1

Under the above restrictions what can you say about $i(x)$?

Exercise 6.8.2

Describe the model setups given in the last section with your choices of i, α , etc.

The idea how to approach finding solutions is the same as in the 1-dimensional case:

- What if u is the product of the initial state i by an exponential decay function of t ?

Let’s test this idea by assuming that it’s true:

$$u(t,x,y)=i(x,y) \cdot g(t).$$

Substituting u into the PDE gives us the following:

$$ig'=K(i_{xx}+i_{yy})g.$$

Let’s rearrange the terms and separate variables:

$$\frac{g'}{Kg}=\frac{i''}{i}.$$

Let’s examine the equation:

$$\frac{g'(t)}{Kg(t)}=\frac{i_{xx}(x,y)+i_{yy}(x,y)}{i(x,y)}.$$

We repeat the argument from the 2-dimensional case: The left-hand side depends only on t is, therefore, constant with respect to x,y , while right-hand side depends only on x,y is, therefore, constant with respect to t . Therefore, this quantity must be constant with respect to both t and x,y ! Both sides are equal to some constant number, say $-\lambda$. We have:

$$g'(t)=-\lambda Kg(t) \quad \text{and} \quad i_{xx}(x,y)+i_{yy}(x,y)=-\lambda i(x,y).$$

The first one is the population ODE (Chapter 1). Its solution is:

$$g(t)=Ce^{-\lambda Kt}.$$

Since we expect exponential decay, we will consider only the case:

$$\lambda>0.$$

Now the PDE:

$$i_{xx}(x,y)+i_{yy}(x,y)=-\lambda i(x,y).$$

We are guessing again: What if u is the sum of two functions that depend only on x and y respectively?
We try:

$$u(x,y) = X(x) + Y(y) \, .$$

We substitute:

$$(X + Y)_{xx} + (X + Y)_{yy} = -\lambda (X + Y) \, .$$

Then:

$$X'' + Y'' = -\lambda (X + Y) \, .$$

We separate the variables again:

$$X'' + \lambda X = -(Y'' + \lambda Y) \, .$$

The argument above applies again and we conclude that this is a constant, say p . We have:

$$X'' + \lambda X = p \quad \text{and} \quad Y'' + \lambda Y = -p \, .$$

The solution is known from Chapter 1:

$$X(x) = A \sin(\sqrt{\lambda} x) + B \cos(\sqrt{\lambda} x) + \frac{p}{\lambda} \, ,$$

and

$$Y(y) = D \sin(\sqrt{\lambda} y) + E \cos(\sqrt{\lambda} y) - \frac{p}{\lambda} \, ,$$

for some constant A, B, D, E .

We also set $p = 0$ and conclude the following.

Theorem 6.8.3: Solutions of Heat PDE Dim 2

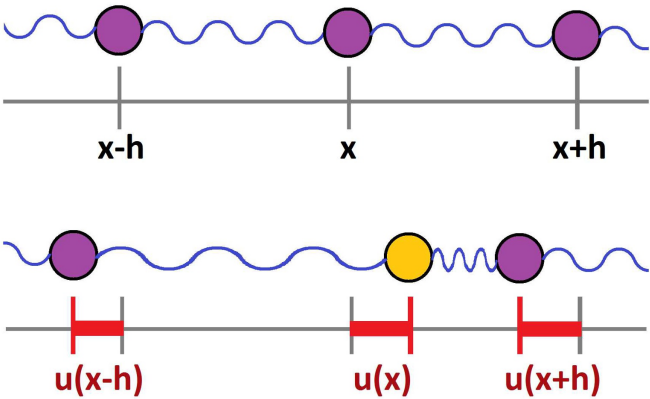
The function $u(t,x,y) =$

$$e^{-\lambda Kt} \big(A \sin(\sqrt{\lambda} x) + B \cos(\sqrt{\lambda} x) + D \sin(\sqrt{\lambda} y) + E \cos(\sqrt{\lambda} y) \big)$$

is a solution to the heat equation for any numbers A, B, D, E .

6.9. Wave propagation in dimension 1: springs and strings

Previously, we studied the motion of an object attached to a wall by a (mass-less) spring. Imagine this time a *string of objects* connected to each other with springs:



Let $u = u(t, x)$ measure the displacement from the equilibrium of the object associated with position x at time t . The cell decompositions for x and t are the same as in the beginning of the chapter.

First, we consider the *spatial variable*, x . Each object is located at a primary node of the cell decomposition with a (possibly variable) distance $h = \Delta x$ to its neighbor and the distance between the secondary nodes is Δs .

According to *Hooke's law*, the force exerted by the spring is:

$$H = -kS,$$

where S is the change of the length of the spring from its equilibrium state and the constant, *stiffness*, k reflects the physical properties of the spring.

Then, if this is the spring that connects locations x_{p-1} and x_p , its compression is the difference of the displacements of the two objects. In other words, we have:

$$S = u(t_j, x_{p-1}) - u(t_j, x_p).$$

Therefore, the force of this spring is

$$H_p = K(t_j, s_p) [u(t_j, x_{p-1}) - u(t_j, x_p)] = -K(t_j, s_p)\Delta_x u(t_j, s_p) = -(k\Delta_x u)(t_j, s_p),$$

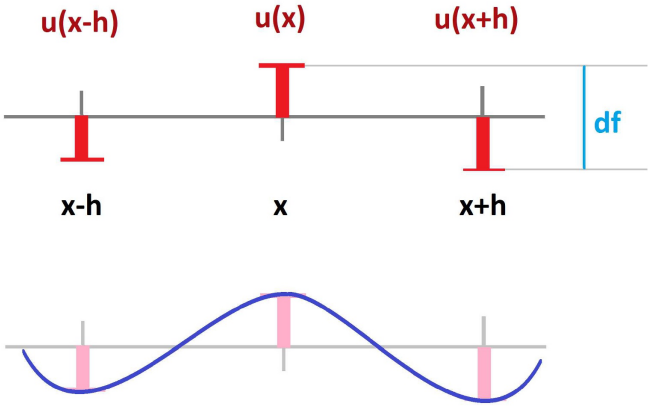
where k is the location- and time-dependent stiffness of the springs.

This formula,

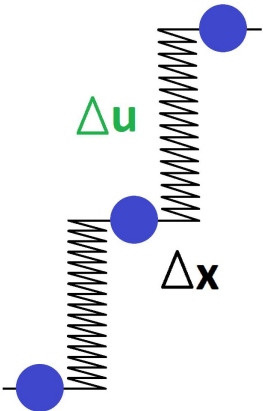
$$H_p = -(k\Delta_x u)(t_j, s_p),$$

expresses the force affecting an object at some location in terms of some quantity that may depend on both location and time. It can have different interpretations for different interpretations of u .

Let's consider an *oscillating string*:



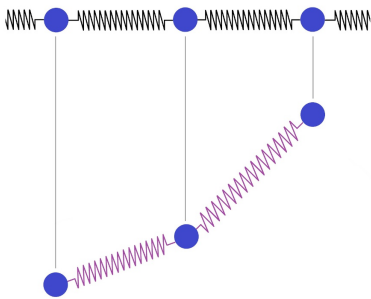
Here, the pieces of the string are vertically displaced (that's $\Delta_x u$) while the waves propagate horizontally. At its simplest, we have a collection of weights that can move only vertically connected by vertical springs:



The horizontal distance between objects, Δx , remain unchanged while the height, $u(t, x)$, varies with time. The vertical displacement is $\Delta_x u$. Applying *Hooke's law* again, we find the exactly the same formula for the force exerted by the spring:

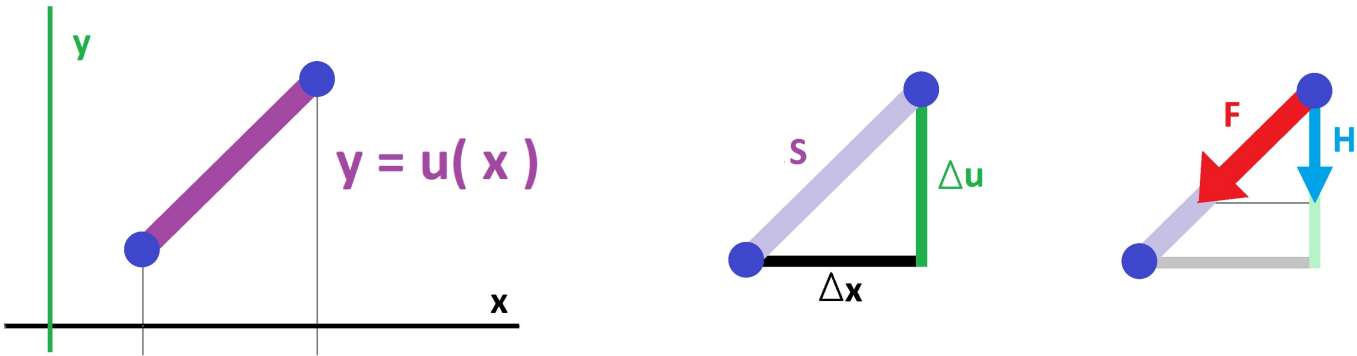
$$H = -k\Delta_x u .$$

A more complex model is the following. We still imagine that the string is made of springs with weights but the spring can go diagonally:



As you can see, the horizontal distance between the weights still remains unchanged; that's Δx . Just as before, we use *Hooke's Law*: The force exerted by the spring is $F = -kS$, where S is the change of the length of the spring from its non-stretched state and k is the stiffness of this spring.

These are the quantities we will have to take into account:



We assume that the equilibrium state of each spring is 0. Then S is simply the distance between the two weights.

Suppose H is the vertical component of the force F . Examining similar triangles reveals:

$$\frac{F}{H} = \frac{S}{\Delta_x u} .$$

Therefore,

$$\frac{-kS}{H} = \frac{S}{\Delta_x u} ,$$

then

$$H = -k\Delta_x u .$$

The equation is identical to the one for the previous approach!

From this equation we derive the differential equation that governs these waves modeled by any of these three approaches.

Let $F(x_i, t_j)$ be the force that acted on the object located at x_i at time t_j . There are two Hooke's forces acting on this object from the two adjacent springs: H_{i-1} and H_i , pulling in the opposite directions. Therefore, we have:

$$\begin{aligned} F(x_i, t_j) &= H_{i-1} && -H_i \\ &= -(k\Delta_x u)(t_j, s_{i-1}) && +(k\Delta_x u)(t_j, s_i) \\ &= \Delta_x (k\Delta_x u)(t_j, x_i) . \end{aligned}$$

Note that this is the same expression as in the right-hand side of the heat equation!

Second, we consider the *temporal variable*, t . Each increment Δt of time may have a different duration.

Now suppose that the object located at x_i has *mass* $m(x_i)$. Then, by the *Second Newton's Law*, the total force is

$$F(t_j, x_i) = m(x_i)a(t_j, x_i),$$

where $a(t_j, x_i)$ is the *acceleration* of object x_i at time t_j . As we know, this is the second difference quotient of the location with respect to time,

$$a(t_j, x_i) = \frac{\Delta^2 u}{\Delta t^2}(t_j, x_i).$$

We have now the *difference wave equation*:

$$\frac{\Delta^2 u}{\Delta t^2}(t_j, x_i) = \frac{1}{m(x_i)}\Delta_x(k\Delta_x u)(t_j, x_i)$$

Now we will derive the recursive formulas in order to simulate wave propagation with a spreadsheet. We make several simplifying assumptions.

First, we assume that the time increments are equal: $\Delta t = 1$. Then the left-hand side of our equation is the following:

$$m\Delta_t^2 u = m(u(t + 1, x) - 2u(t, x) + u(t - 1, x)).$$

For the right-hand side, we can use the original expression:

$$\Delta_x(k\Delta_x u) = k[u(t, x - 1) - u(t, x)] + k[u(t, x + 1) - u(t, x)].$$

Second, we assume that k and m are constant. Then just solve for $u(x, t + 1)$:

$$u(t + 1, x) = 2u(t, x) - u(t - 1, x) + \alpha(u(t, x + 1) - 2u(t, x) + u(t, x - 1)),$$

where

$$\alpha = \frac{k}{m}.$$

To visualize the formula, we arrange the terms in a table to be implemented as a spreadsheet:

	$x - 1$	x	$x + 1$
$t - 1$	$-u(t - 1, x)$		
t	$= \alpha u(t, x - 1)$	$+ 2(1 - \alpha)u(t, x)$	$+ \alpha u(t, x + 1)$
$t + 1$	$u(t + 1, x)$		

Even though the right-hand side is the same, the table is different from that of the heat equation. The presence of the second derivative with respect to time makes it necessary to look *two steps back*, not just one. That's why we also have *two initial conditions*.

We suppose, for simplicity, that $\alpha = 1$.

Example 6.9.1: algebra

Choosing the simplified settings allows us to easily solve the following initial value problem by hand:

$$\Delta_t^2 u = \Delta_x^2 u;$$
$$u(t, x) = \begin{cases} 1 & \text{if } t = 0, x = 1; \\ 0 & \text{if } t = 0, x \neq 1; \\ 1 & \text{if } t = 1, x = 2; \\ 0 & \text{if } t = 1, x \neq 2. \end{cases}$$

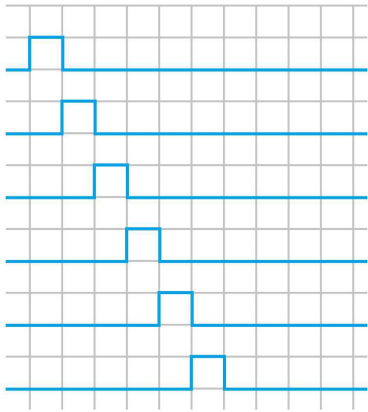
Initially, the wave has a single bump and then the bump moves one step from left to right. The

negative values of x are ignored.

Now, setting $k = 1$ makes the middle term in the table disappear. Then every new term is computed by taking an alternating sum of the three terms above, as shown below:

$t \backslash x$	1	2	3	4	5	6	7	..
0	1	0	0	0	0	[0]	0	..
1	0	1	0	0	[0]	0	[0]	..
2	0	0	1	0	0	(0)	0	..
3	0	0	0	1	0	0	0	..
4	0	0	0	0	1	0	0	..
5	0	0	0	0	0	1	0	..
..

We can see that the wave is a single bump running from left to right at speed 1:



Exercise 6.9.2

Set up and solve an IVP with 2 bumps, n bumps.

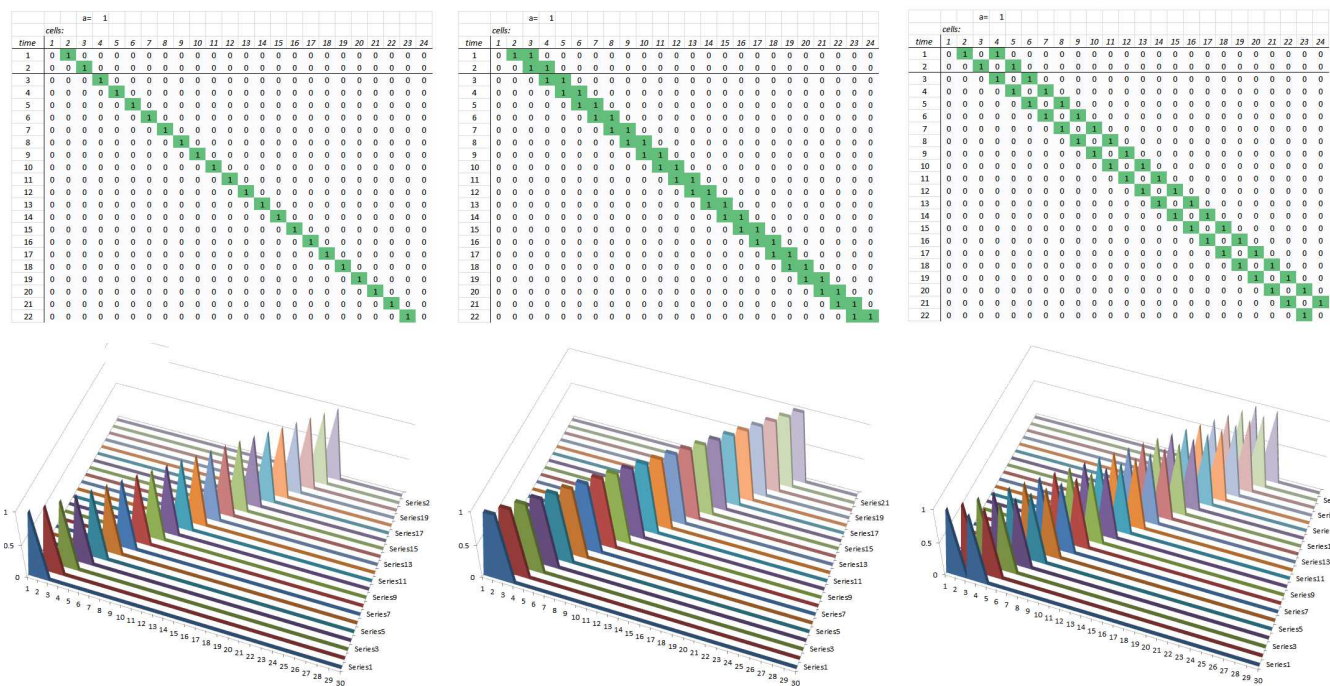
The simplest way to implement this dynamics with a spreadsheet is to use the first two rows for the initial conditions and then add one row for every moment of time. The Excel formula is:

$$=R1C5*R[-1]C[-1]+2*(1-R1C5)*R[-1]C+R1C5*R[-1]C[1]-R[-2]C$$

Here cell **R1C5** contains the value of α .

Example 6.9.3: bumps

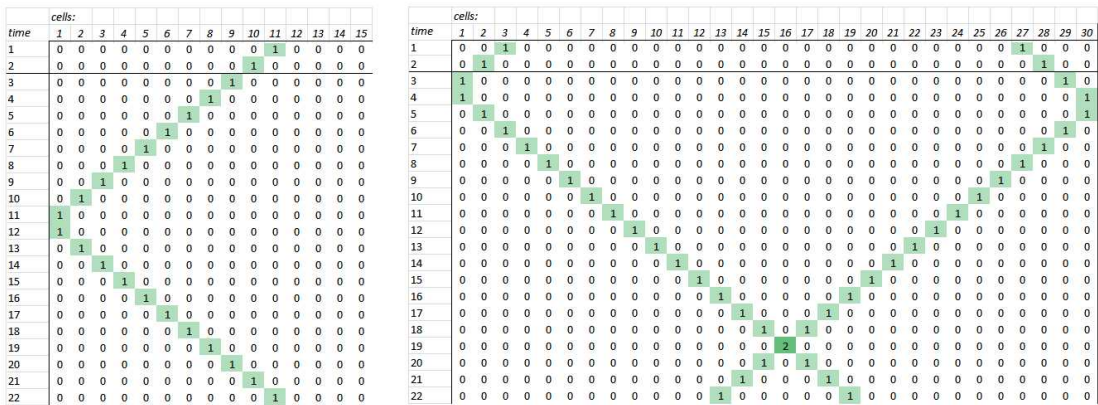
The simplest propagation pattern is given by $\alpha = 1$. Below we show the propagation of a single bump, a two-cell bump, and two bumps:



In the second row, the swing of the wave is visualized as a function of two variables.

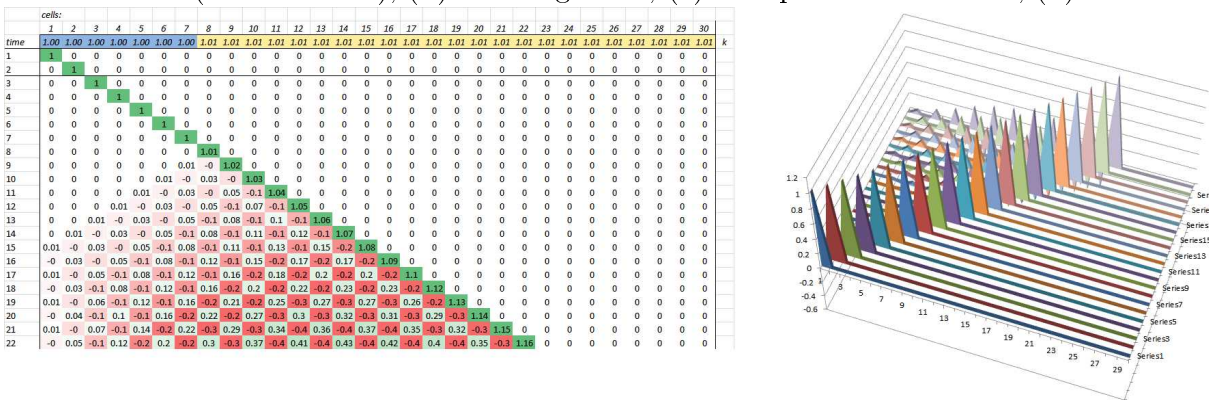
Exercise 6.9.4

Modify the spreadsheet to introduce walls (one and then two) into the picture:



Exercise 6.9.5

Modify the spreadsheet to accommodate non-constant data by adding variability to the following: (a) the stiffness k (shown below), (b) the weights m , (c) the space intervals Δx , (d) the time intervals Δt .



Exercise 6.9.6

Implement a spreadsheet simulation for the case of non-constant m . Hint: You will need *two* buffers.

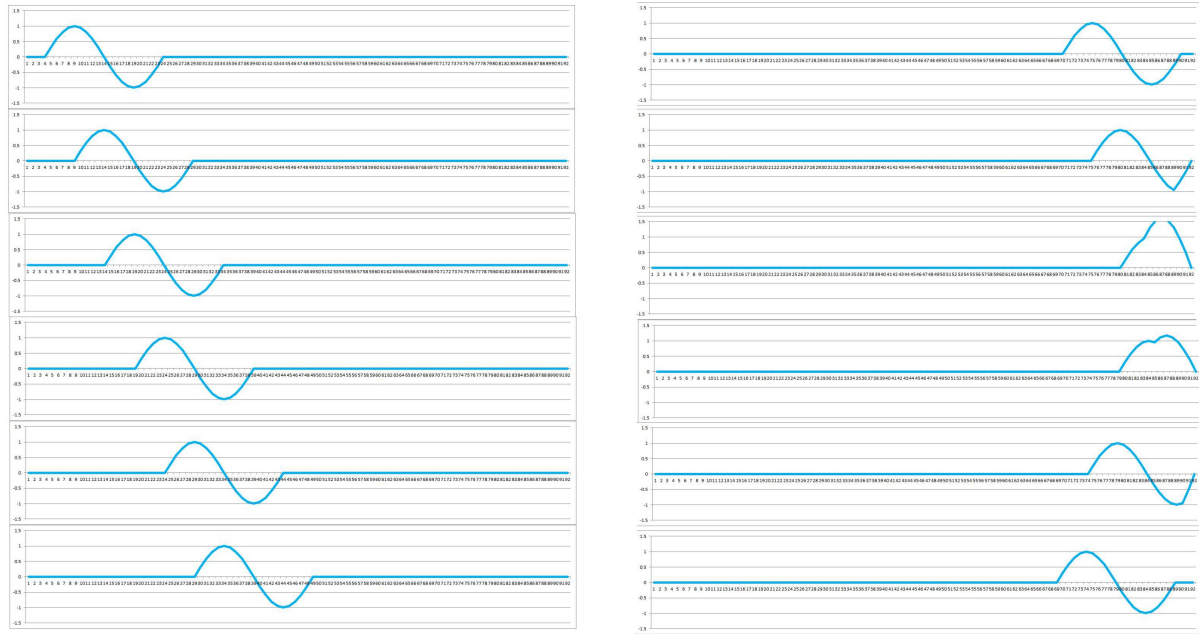
Below we consider a rope with fixed ends. This is reflected in the boundary conditions:

$$u(t,0) = u(t,n) = 0, \text{ for all } t.$$

We again choose $\alpha = 1$.

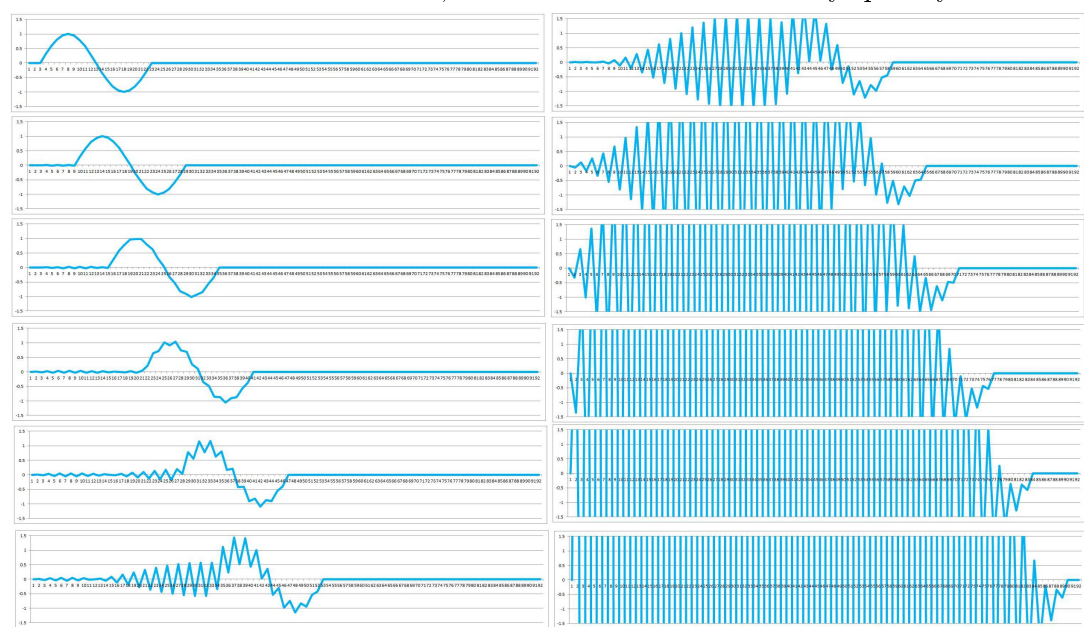
Example 6.9.7: whip

The simulation starts with an arbitrary shape given by the first initial condition $u(0,x)$. It is a small piece of a sinusoid. The second initial condition is just a shift, $u(1,x) = u(0,x - 1)$, of the first. The simulation continues to produce this shift at every iteration; in fact, the magnitude of the initial shift is the speed of propagation of the shape. The rope behaves like a whip:



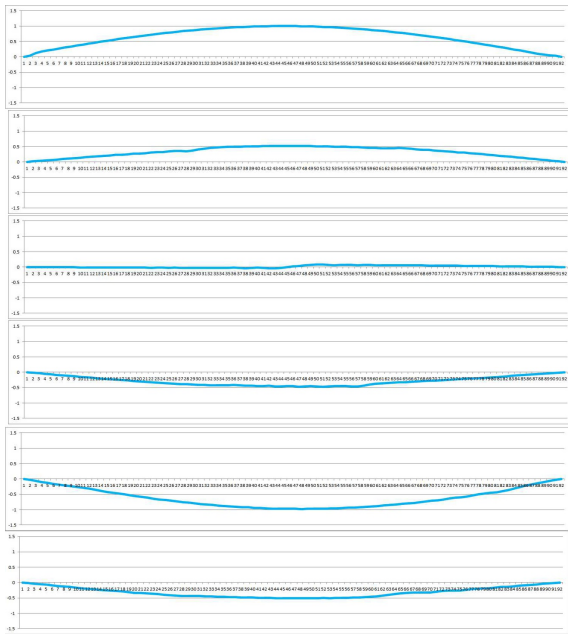
This effect is known as a “traveling wave”. Furthermore, because the end is attached to the wall, the wave bounces off the end.

When the coefficient α is 1.01 instead of 1, the model breaks down very quickly:

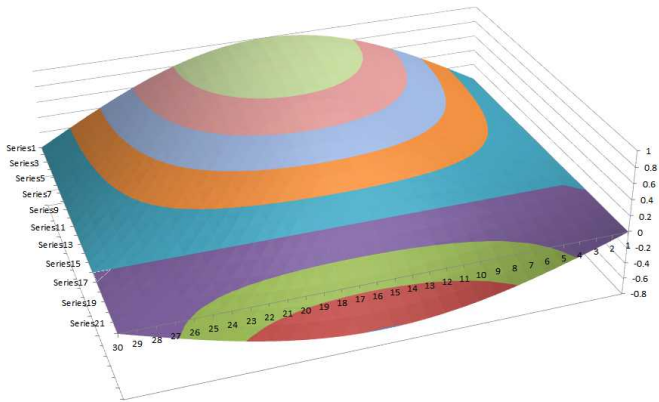


Example 6.9.8: vibrating string

The simulation starts with a sinusoid given by the first initial condition $u(x,0) = \sin(\pi x/n)$. The second initial condition is an exact copy, $u(1,x) = u(0,x)$, of the first. The initial velocity, then, is zero (there is of course acceleration). The simulation produces more sinusoids and a perfect vibration:



We can also visualize u by its graph as a function of two variables:



Exercise 6.9.9

What if half of the string is made of a stiffer material?

6.10. The wave PDE

We now consider the *continuous case* of wave propagation.

Suppose the function u is defined for *all* x and t within some open subset U of the plane and it is sampled at the nodes of a cell decomposition of rectangle $[0, b] \times [0, d]$ contained in that subset. We now utilize the results from the last section in the special case of a *finite* string of weights and springs:

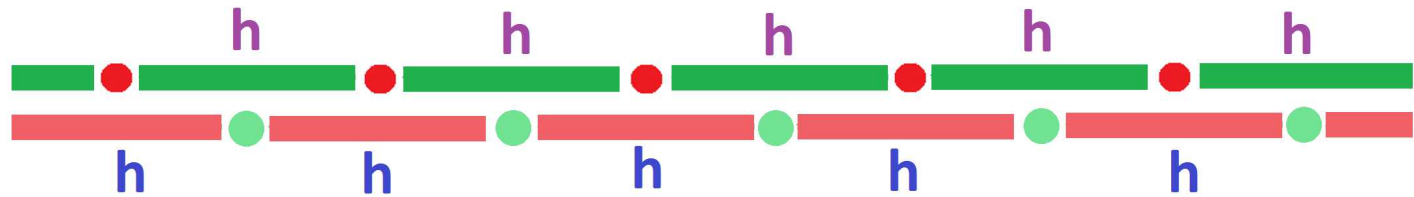


These are our assumptions:

- The array of N identical weights; m is constant.
- The weights are distributed evenly over the length $L = N\Delta x$ of the string.

- The total mass is $M = Nm$.
- The springs are identical; k is constant.
- The *total spring constant* of the array is $K = k/N$.
- The cell decomposition of the string is uniform: $\Delta x = \Delta s = h$.
- The cell decomposition of the time interval is also uniform: $\Delta t = \Delta q$.

We will need again a dual decomposition of the interval:



We can rewrite the difference equation for wave propagation from the last section as follows:

$$m \frac{u(t + \Delta t, x) - 2u(t, x) + u(t - \Delta t, x)}{(\Delta t)^2} = k [u(t, x + \Delta x) - 2u(t, x) + u(t, x - \Delta x)] .$$

Furthermore:

$$\begin{aligned} \frac{u(t + \Delta t, x) - 2u(t, x) + u(t - \Delta t, x)}{(\Delta t)^2} &= \frac{K(\Delta x)^2}{m} \frac{u(t, x + \Delta x) - 2u(t, x) + u(t, x - \Delta x)}{(\Delta x)^2} \\ &= \frac{k/N(N\Delta x)^2}{Nm} \frac{u(t, x + \Delta x) - 2u(t, x) + u(t, x - \Delta x)}{(\Delta x)^2} \\ &= \frac{KL^2}{M} \frac{u(t, x + \Delta x) - 2u(t, x) + u(t, x - \Delta x)}{(\Delta x)^2} . \end{aligned}$$

Solving for $u(t + \Delta t, x)$, we obtain the recursive formula:

$$u(t + \Delta t, x) = 2u(t, x) - u(t - \Delta t, x) + \alpha \left(\frac{\Delta t}{\Delta x} \right)^2 [u(t, x + \Delta x) - 2u(t, x) + u(t, x - \Delta x)] ,$$

where

$$\alpha = \frac{KL^2}{M} .$$

Exercise 6.10.1

Set up and solve the IVP for the finite string, for the simplified settings. Hint: mind the ends.

We have also found the *discrete wave equation*:

$$\frac{\Delta^2 u}{\Delta t^2} = \alpha \frac{\Delta^2 u}{\Delta x^2} .$$

It is a PDE with respect to second difference quotients.

We refine our cell decomposition of the rectangle and take the limit of this equation as $\Delta x \rightarrow 0$ and $\Delta t \rightarrow 0$. We have the *wave equation*:

$$\frac{\partial^2 u}{\partial t^2} = \alpha \frac{\partial^2 u}{\partial x^2}$$

It is a PDE with respect to second derivatives.

A solution u of this PDE is a function:

- defined and continuous on the rectangle,
- twice differentiable with respect to t inside of it, and
- twice differentiable with respect to x inside the rectangle, such that
- the equation is satisfied for each pair (x, t) such that x is in $(0, b)$ and t is in $(0, d)$.

In contrast to the heat transfer PDE, we will be able to find solutions for *all* possible initial conditions. In order to solve this PDE, we introduce new variables:

$$p = x + \alpha t \quad \text{and} \quad q = x - \alpha t,$$

Then,

$$x = \frac{1}{2}(p + q) \quad \text{and} \quad t = \frac{1}{2\alpha}(p - q).$$

Exercise 6.10.2

What is the matrix of this *linear* transformation?

The *new unknown function* is the result of the substitution:

$$v(p, q) = u(t, x) = u\left(\frac{1}{2}(p + q), \frac{1}{2\alpha}(p - q)\right),$$

and

$$u(t, x) = v(x + \alpha t, x - \alpha t).$$

Theorem 6.10.3: Mixed Second Derivative of Wave

Suppose u is twice continuously differentiable and satisfies the wave equation. Then the mixed second partial derivative of v is zero.

Proof.

Let's first list the partial derivatives of the old variables with respect to the new ones:

$$\frac{\partial x}{\partial p} = \frac{1}{2}, \quad \frac{\partial x}{\partial q} = \frac{1}{2} \quad \text{and} \quad \frac{\partial t}{\partial p} = \frac{1}{2\alpha}, \quad \frac{\partial t}{\partial q} = -\frac{1}{2\alpha}.$$

Now let's differentiate v . We use the *Chain Rule* (Chapter 4HD-3) several times. First:

$$\frac{\partial v}{\partial p}(p, q) = \frac{\partial}{\partial p}u(t, x) = \frac{\partial u}{\partial x} \frac{\partial x}{\partial p} + \frac{\partial u}{\partial t} \frac{\partial t}{\partial p} = \frac{\partial u}{\partial x} \frac{1}{2} + \frac{\partial u}{\partial t} \frac{1}{2\alpha}.$$

Second:

$$\begin{aligned} \frac{\partial^2 v}{\partial q \partial p}(p, q) &= \frac{\partial}{\partial q} \left(\frac{\partial v}{\partial p}(p, q) \right) = \frac{\partial}{\partial q} \left(\frac{\partial u}{\partial x} \frac{1}{2} + \frac{\partial u}{\partial t} \frac{1}{2\alpha} \right) \quad \text{Substitute.} \\ &= \left(\frac{\partial^2 u}{\partial x^2} \frac{\partial x}{\partial q} + \frac{\partial^2 u}{\partial t \partial x} \frac{\partial t}{\partial q} \right) \frac{1}{2} + \left(\frac{\partial^2 u}{\partial t \partial x} \frac{\partial x}{\partial q} + \frac{\partial^2 u}{\partial t^2} \frac{\partial t}{\partial q} \right) \frac{1}{2\alpha} \\ &= \left(\frac{\partial^2 u}{\partial x^2} \frac{1}{2} + \frac{\partial^2 u}{\partial x \partial t} \left(-\frac{1}{2\alpha} \right) \right) \frac{1}{2} + \left(\frac{\partial^2 u}{\partial t \partial x} \frac{1}{2} + \frac{\partial^2 u}{\partial t^2} \left(-\frac{1}{2\alpha} \right) \right) \frac{1}{2\alpha} \quad \text{By Clairaut's theorem...} \\ &= \frac{\partial^2 u}{\partial x^2} \frac{1}{2} \frac{1}{2} - \frac{\partial^2 u}{\partial t^2} \frac{1}{2\alpha} \frac{1}{2\alpha}, \quad \dots \text{ and because } u \text{ satisfies the wave equation.} \\ &= 0. \end{aligned}$$

Solving the resulting PDE,

$$\frac{\partial^2 v}{\partial p \partial q} = 0,$$

is easy. Indeed, any function that depends on only on p satisfies it:

$$v(p, q) = f(p).$$

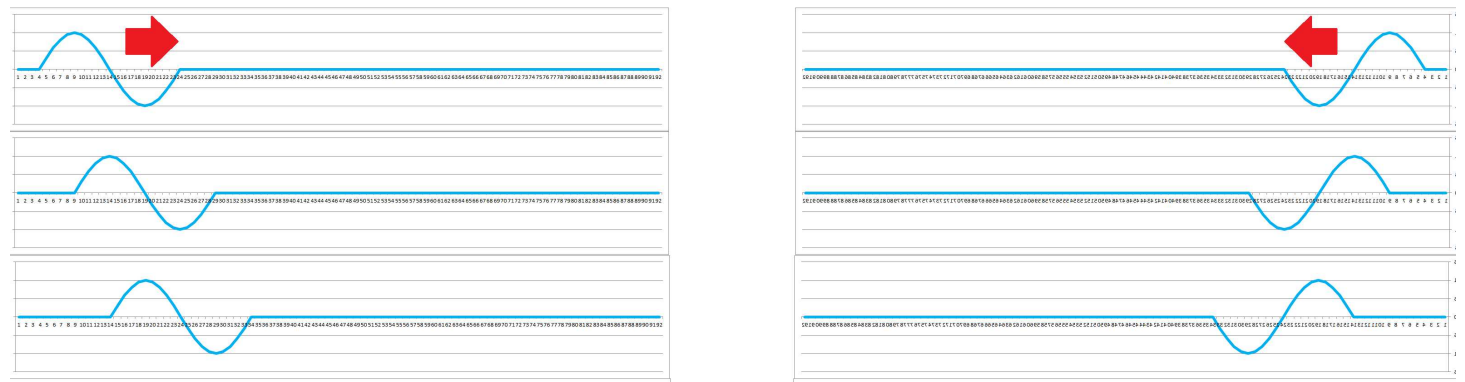
And so does any function that depends only on q :

$$v(p, q) = g(q).$$

We have then two solutions of the original PDE:

$$u(t, x) = f(x + \alpha t) \quad \text{and} \quad u(t, x) = g(x - \alpha t),$$

for any choice of f and g . What are these solutions? They are the *traveling waves* that we saw in the last section: the shape remains the same as it is moving along the rope at the speed α . The two above are a left traveling wave and a right traveling wave respectively:



The only difference is that this time the functions have to be differentiable.

Furthermore, the sum of two left/right traveling waves is a left/right traveling wave. Moreover, the sum of a left traveling wave and a right traveling wave is also a solution! Indeed, the mixed second derivative of

$$v(p, q) = f(p) + g(q),$$

for any pair of *arbitrary* functions f and g , is zero. Therefore, we have the following.

Theorem 6.10.4: Solution of Wave Equation

For any two twice differentiable functions f and g , the function

$$u(t, x) = f(x + \alpha t) + g(x - \alpha t)$$

is a solution of the wave equation.

Exercise 6.10.5

Have we found *all* solutions?

Example 6.10.6: standing waves

What if the two traveling waves are identical? What if f and g are the same function? Let's choose:

$$f(p) = g(p) = \sin(p).$$

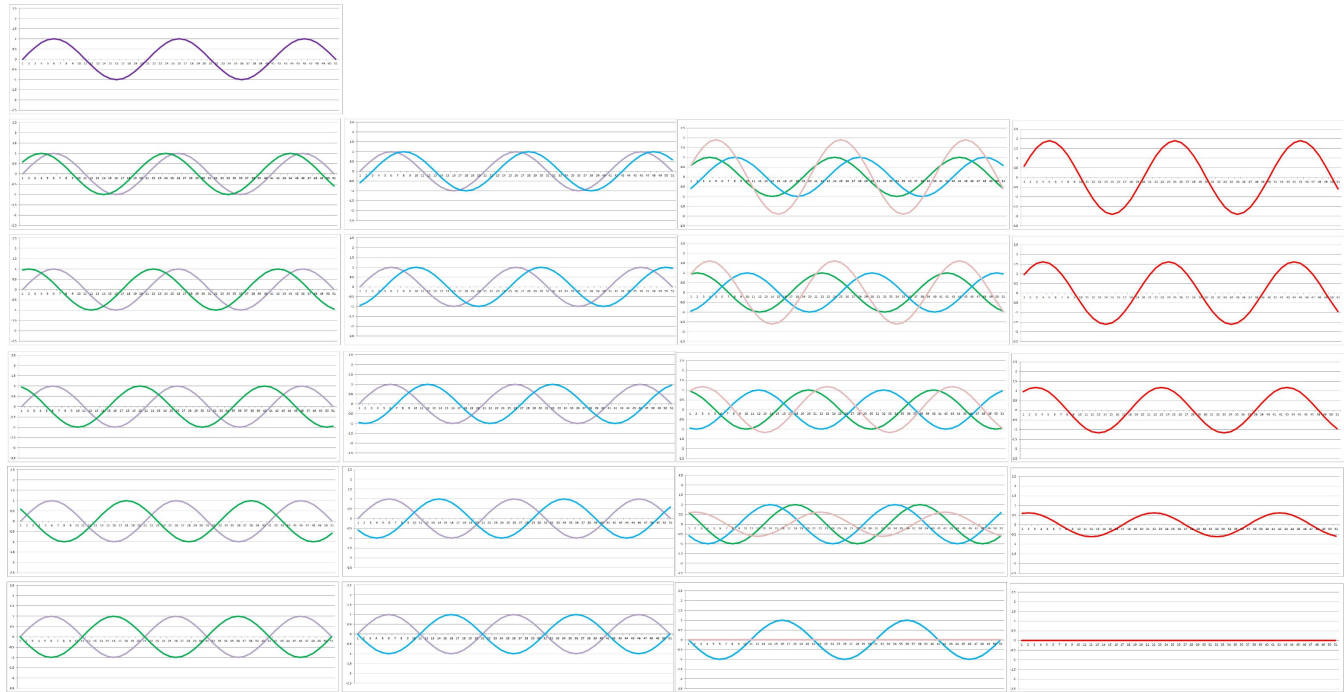
Then we have two traveling waves, the right,

$$u_1(t, x) = \sin(x - \alpha t),$$

and left,

$$u_2(t,x) = \sin(x + \alpha t) .$$

What is their sum u ?



There seems to be no traveling! Let’s confirm this algebraically. We have:

$$u(t,x) = u_1(t,x) + u_2(t,x) = \sin(x - \alpha t) + \sin(x + \alpha t) = 2 \sin x \cos(\alpha t) ,$$

by the sum-to-product trigonometric identity,

$$\sin a + \sin b = 2 \sin \left(\frac{a + b}{2} \right) \cos \left(\frac{a - b}{2} \right) .$$

Here the sinusoid of $u = 2 \sin x$ is stretched vertically by a time-dependent multiple, $\cos(\alpha t)$. This describes a wave that oscillates – up and down – but doesn’t move horizontally. An oscillating spring is an example of this that we saw in the last section.

Next, we impose initial conditions on this function: one for the (vertical) location of each weight on the string and one for the (vertical) velocity. We have the following theorem.

Theorem 6.10.7: D’Alembert’s Formula

The initial value problem of the wave equation,

$$\begin{cases} \frac{\partial^2 u}{\partial t^2} = \alpha^2 \frac{\partial^2 u}{\partial x^2} \\ u(0,x) = h_0(x) \\ \frac{\partial u}{\partial t}(0,x) = h_1(x) \end{cases}$$

has a solution:

$$u(t,x) = \frac{h_0(x - \alpha t) + h_0(x + \alpha t)}{2} + \frac{1}{2\alpha} \int_{x-\alpha t}^{x+\alpha t} h_1(s) \, ds$$

when

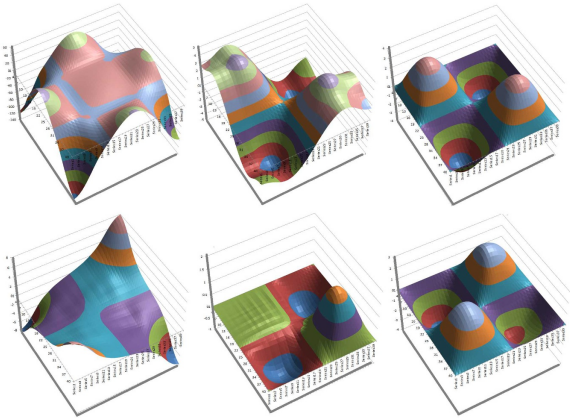
- h_0 is twice continuously differentiable, and
- h_1 is continuously differentiable.

Exercise 6.10.8

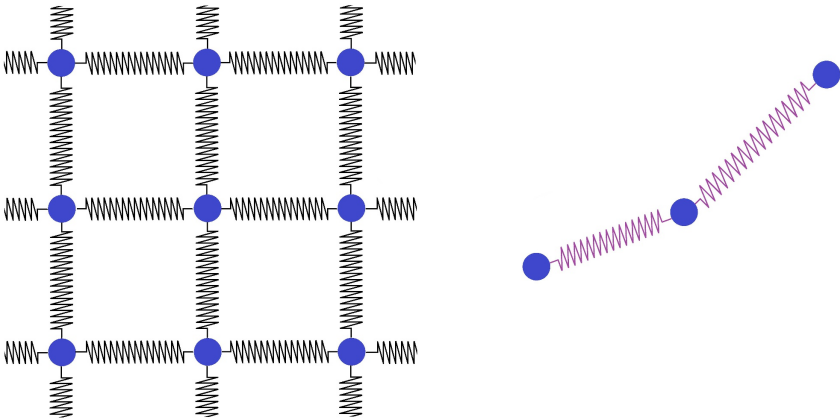
Confirm the formula. Hint: Split the integral into two.

6.11. Wave propagation in dimension 2: a membrane

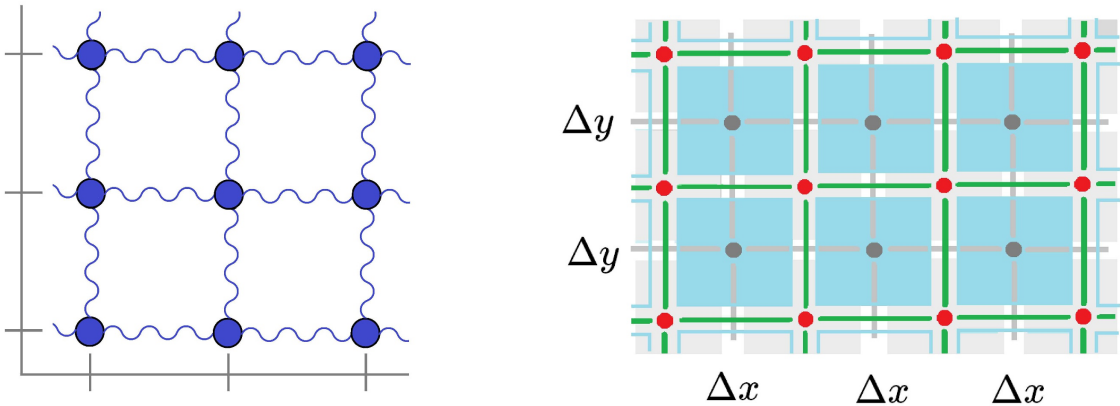
We model a membrane of a drum as well as the surface of a liquid:



Here, the pieces of the membrane are vertically displaced and the waves propagate horizontally. Just as before, we represent this membrane by weights connected by springs but, this time it is not a string but an *array*:



On the left, we see the view from above and on the right from aside. We use a cell decomposition of this rectangle and this is the domain of u , which is a 0-form.



The cell decomposition is given by the lengths of the edges: Δx and Δy .

Recall that a cell decomposition of a *box* B in the xyt -space comes from cell decompositions of its three edges as described in [Chapter 4HD-5](#) and earlier in this chapter:

$$\begin{array}{cccccc} t_0 & q_1 & t_1 & q_2 & t_2 & q_3 & \dots \\ x_0 & s_1 & x_1 & s_2 & x_2 & s_3 & \dots \\ y_0 & p_1 & y_1 & p_2 & y_2 & p_3 & \dots \end{array}$$

We make a simplifying assumption that all weights are equal. We use the analysis of the 1-dimensional case: if H is the vertical component of the force F of the spring, then

$$H = -K\Delta_x u \quad \text{or} \quad H = -K\Delta_y u,$$

depending on whether this spring is aligned with the x - or the y -axis.

The forces exerted on the object at location x are the *four* forces of the four springs attached to it. Each term is the difference: two with respect to x and two with respect to y . The algebra that follows is identical to that we used for the heat equation in dimension 2.

For the weight located at (x_i, y_j) , this is the *total force* at time t_k :

$$F(t_k, x_i, y_j) = \left(\begin{array}{ccc} & \bullet & \\ -K(t_k, s_{i-1}, y_j) \Delta_x u(t_k, s_{i-1}, y_j) & -K(t_k, s_i, y_{j-1}) \Delta_y u(t_k, s_i, p_{j-1}) & \bullet \\ & \bullet & +K(t_k, s_i, y_j) \Delta_x u(t_k, s_i, y_j) \end{array} \right) + K(t_k, s_i, y_j) \Delta_y u(t_k, s_i, p_j)$$

The four terms are the forces of the four springs and they are arranged accordingly.

The *difference equation* for the (vertical) displacements of the weights $u = u(t, x, y)$ comes from the *Second Newton's Law* (mass times acceleration is force):

$$m(x_i, y_j) \frac{\Delta^2 u}{\Delta t}(t_k, x_i, y_j) = F(t_k, x_i, y_j).$$

Since the acceleration is now known:

$$a(t_k, x_i, y_j) = \frac{1}{m(x_i, y_j)} F(t_k, x_i, y_j),$$

we have the two *recursive formulas* for the velocity and the location derived just as before:

$$\begin{aligned} v(q_k, x_i, y_j) &= v(q_{k-1}, x_i, y_j) + a(t_k, x_i, y_j) \Delta t_k, \\ u(t_k, x_i, y_j) &= u(t_{k-1}, x_i, y_j) + v(q_k, x_i, y_j) \Delta t_k. \end{aligned}$$

The spreadsheet consists of several sheets computed consecutively:

- the stiffness for each spring,
- the mass of each weight,
- the initial location for each weight,
- the next location for each weight (i.e., the velocity),
- the first buffer (copied from the second buffer),
- the second buffer (copied from the current values),
- the difference of locations of the ends and the Hook's force for each spring,
- the total force for each weight,
- the current velocity, and finally
- the current location for each weight.

Two examples are as follows. The domain is chosen to be the square $[1, 40] \times [1, 40]$ with $\Delta x = \Delta y = 1$.

Example 6.11.1: drum

A drum is a circular membrane the edge of which is attached to a ring. In other words, the boundary condition is

$$u(t,x,y)=0 \text{ when } x^2+y^2>20^2.$$

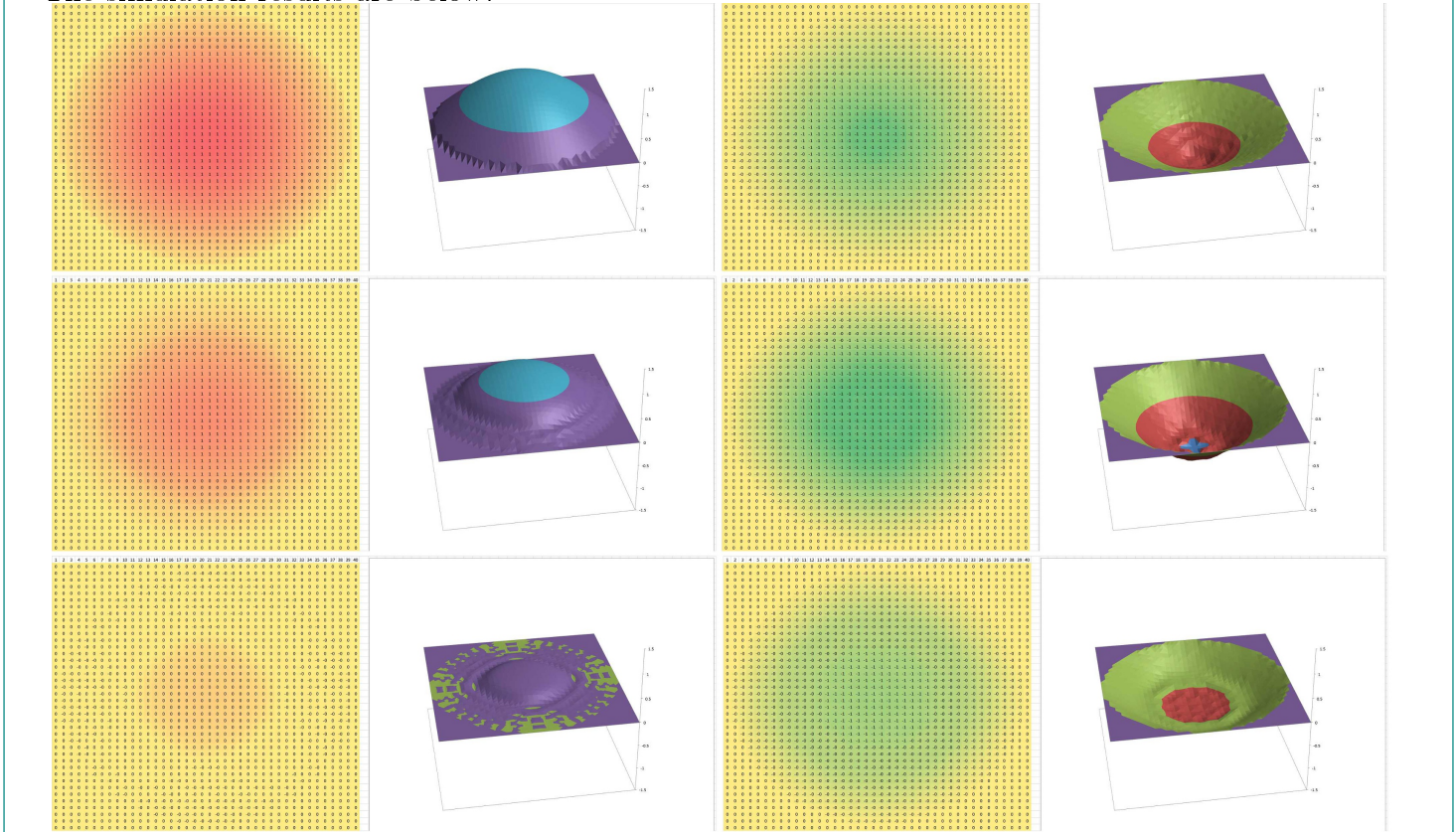
We choose the initial shape of the drum to be a rotated sinusoid:

$$u(0,x,y)=\cos\left(\sqrt{(x-20)^2+(y-20)^2}\frac{\pi}{40}\right),$$

and zero initial velocity

$$u(1,x,y)=u(0,x,y).$$

The simulation results are below:



The drum skin shrinks to flat and then creates the same shape on the other side.

Example 6.11.2: breakwater

Seemingly having nothing to do with it, can our model reasonably reproduce waves of the surface of a liquid? Below we show a simulation of a breakwater protecting a harbor. It starts with a single “pulse” (tsunami?) outside the harbor,

$$u(0,20,3)=1, \; u(0,x,y)=0 \text{ for the rest of } (x,y),$$

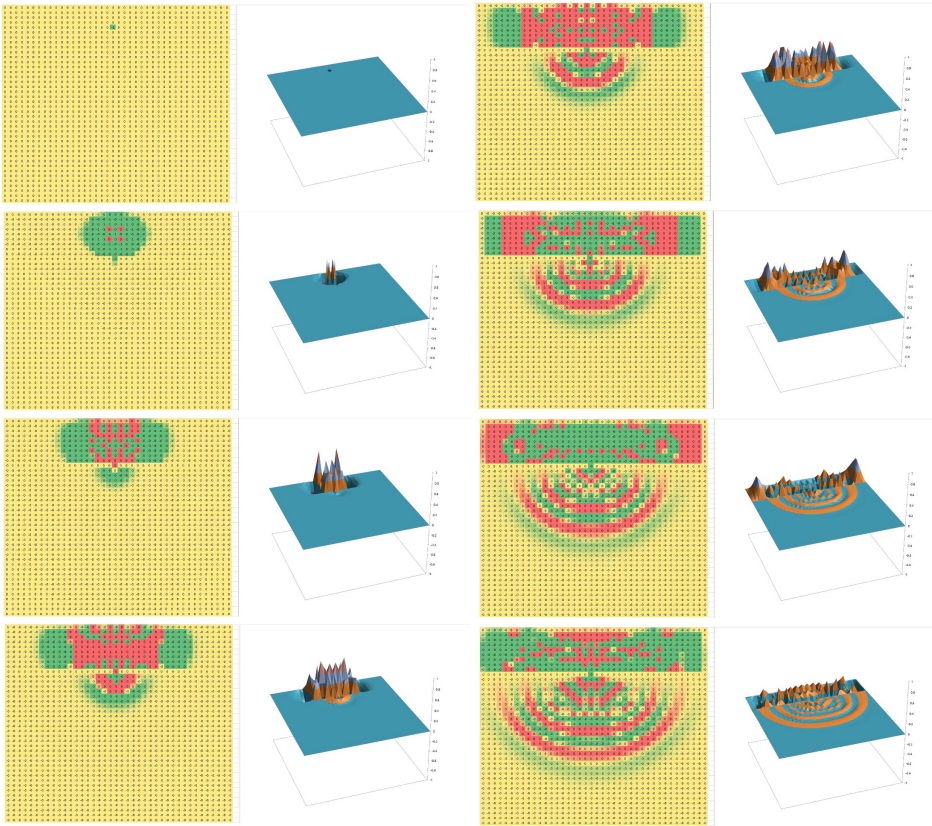
and

$$u(1,x,y)=0 \text{ for all } (x,y).$$

The breakwater is represented by a row of fixed values:

$$u(t,x,10)=0 \text{ for all } x\neq 20,$$

with a single gap. The simulation results are below:



The large waves outside produce very mild, circular waves inside.

We repeat the algebra we used in our analysis of heat transfer for dimension 2. When K is constant, as another simplifying assumption, the total force is simplified:

$$F(x_i, y_j, t_k) =$$
$$= K \begin{pmatrix} \bullet & -\Delta_y u(t_k, s_i, p_{j-1}) & \bullet \\ -\Delta_x u(t_k, s_{i-1}, y_j) & & +\Delta_x u(t_k, s_i, y_j) \\ \bullet & +\Delta_y u(t_k, s_i, p_j) & \bullet \end{pmatrix} = K (\Delta_x^2 u(t_k, x_i, y_j) + \Delta_y^2 u(t_k, x_i, y_j)) .$$

These are the second differences. Our *partial difference equation* becomes:

$$\frac{\Delta^2 u}{\Delta t^2} = \frac{K}{m} (\Delta_x^2 u + \Delta_y^2 u) ,$$

where $\alpha > 0$ is a constant.

Exercise 6.11.3

Derive a version of this equation for a variable K .

Just as in the 1-dimensional case, we can argue the following two points. First, a longer spring is less stiff than a shorter one made of the same material and, therefore, K is inversely proportional to $\Delta x = \Delta y$. Second, the weights are, in fact, the weights of the springs and, therefore, m is proportional to Δx . Then our difference equation becomes:

$$\frac{\Delta^2 u}{\Delta t^2} = \frac{\alpha}{\Delta x^2} (\Delta_x^2 u + \Delta_y^2 u) = \alpha \left(\frac{\Delta_x^2 u}{\Delta x^2} + \frac{\Delta_y^2 u}{\Delta y^2} \right) .$$

Furthermore, when u is defined throughout the region, the difference equation corresponds to the *wave equation* for partial derivatives:

$$\frac{\partial^2 u}{\partial t^2} = \alpha \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) ,$$

with the last expression is, again, the *Laplace operator* of u .

Exercise 6.11.4

Derive this equation from the last.

Exercises

Contents

1 Exercises: Basics	410
2 Exercises: Analytical methods	412
3 Exercises: Euler’s method	413
4 Exercises: Generalities	414
5 Exercises: Models and setting up ODEs	415
6 Exercises: Qualitative analysis	416
7 Exercises: Systems	417
8 Exercises: Second order	418
9 Exercises: Advanced	419
10 Exercises: PDEs	420
11 Exercises: Computing	421

1. Exercises: Basics

Exercise 1.1

Find *all* antiderivatives of x^{-2} defined on $(-\infty, 0) \cup (0, +\infty)$.

Exercise 1.3

Verify that the function $y = cx^2$ is a solution of the differential equation:

$$xy' = 2y.$$

Are there any others?

Exercise 1.2

Find the eigenvalues and the eigenvectors of the following matrix:

$$F = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

Exercise 1.4

Find all curves perpendicular to the family of curves:

$$xy^2 = C.$$

Exercise 1.5

Find all curves perpendicular to the family of curves:
$$x^2y = C.$$

Exercise 1.6

Find all curves perpendicular to the family of curves:
$$xy^2 = C.$$

Exercise 1.7

Verify that $-2x^2y + y^2 = 1$ is an implicit solution of the differential equation $2xy + (x^2 - y)y' = 0$. Find one explicit solution.

Exercise 1.8

In the square $[-3, 3] \times [-3, 3]$, plot the direction field (slope field) for the differential equation $y \frac{dy}{dx} = -x$. Sketch the stationary solution and 3 other solution curves.

Exercise 1.9

Solve the differential equation: $y' + 2xy^2 = 0$.

Exercise 1.10

Solve the IVP:
$$(4y + 2x - 5)dx + (6y + 4x - 1)dy, \quad y(-1) = 2.$$

Exercise 1.11

Your location is given below as a function of time. Find the acceleration.

t	0.0	0.5	1.0	1.5	2.0	2.5	3.0	3.5
y	2	5	10	5	0	0	3	7

2. Exercises: Analytical methods

Exercise 2.1

Solve the differential equation by using an appropriate substitution: $y' = 1 + e^{y-x+5}$.

Exercise 2.2

For the differential equation $y' = 1 + e^{3y+5}$, execute the substitution $u = 3y + 5$.

Exercise 2.3

Solve by separating variables:

$$y' = xy.$$

Exercise 2.4

Solve by the method of integrating factor:

$$y' = y/x.$$

Exercise 2.5

Solve the differential equation:

$$y' + 2xy^2 = 0.$$

Exercise 2.6

Solve the differential equation:

$$x \frac{dy}{dx} - y = x^2 \sin x.$$

Exercise 2.7

Solve the IVP:

$$(4y + 2x - 5) dx + (6y + 4x - 1) dy = 0, \quad y(-1) = 2.$$

Exercise 2.8

Solve the differential equation by using an appropriate substitution:

$$\frac{dy}{dx} = 1 + e^{y-x+5}.$$

Exercise 2.9

Solve the differential equation by using an appropriate substitution:

$$\frac{dy}{dx} = 2 + \sqrt{y - 2x + 3}.$$

Exercise 2.10

Solve the differential equation

$$xyy' = x^2 + 3y^2.$$

Exercise 2.11

Solve the differential equation

$$(2x^2y + 3x^2) dx + (2x^2y + 4y^3) dy.$$

Exercise 2.12

Solve the differential equation:

$$y' = -6xy.$$

3. Exercises: Euler’s method

Exercise 3.1

Carry out $n = 4$ steps of Euler’s method with $h = .5$ for the following initial value problem:

$$y' = y - x, \; y(0) = 2.$$

Exercise 3.2

Suppose y is the solution of the initial value problem

$$y' = x^2 + y^2, \; y(0) = 0.$$

Find $y(1)$ by means of Euler’s method with step $h = .2$.

Exercise 3.3

Use Euler’s method with $n = 4$ steps to estimate the solution of the initial values problem:

$$y' = 2x - y, \; y(0) = 1,$$

on the interval $[0, 1]$.

Exercise 3.4

Use Euler’s method with 4 steps to estimate the solution of the initial values problem:

$$y' = 2x - y, \; y(0) = 1,$$

on the interval $[0, 1]$.

Exercise 3.5

Linearize the following ODE at $x = 2$:

$$y' = \sqrt{x - 1} - 1.$$

Do not solve.

Exercise 3.6

Set up (don’t solve) initial values problems – for x and y – for the following situation: An object is thrown up from a building of height h at 45 degrees with speed s .

Exercise 3.7

Solve by separating the variables:

$$y' = xy.$$

4. Exercises: Generalities

Exercise 4.1

Provide the definition of the uniqueness property and sketch an example of a solution set without it.

Exercise 4.2

Indicate if the following statements are true or false.

- 1. $y' = y^2$ is a second order ODE.
- 2. $y(t) = |t|$ is a solution of the IVP $y' = 1, y(0) = 0$.
- 3. $y(t) = -t$ is a solution of a DE with the slope field below.
- 4. The IVP $y' = \frac{1}{t^2 + 1}, y(0) = 1$ has more than one solution.
- 5. The ODE $y' = y t^2 \sin t$ can be solved by separation of variables.

Exercise 4.3

Verify that $-2x^2y + y^2 = 1$ is an implicit solution of the differential equation

$$2xy + (x^2 - y)\frac{dy}{dx} = 0.$$

Find one explicit solution.

Exercise 4.4

(a) State the Existence-Uniqueness Theorem for a system of two linear equations. (b) For what values t_0, a, b does the theorem guarantee existence and uniqueness of the IVP

$$x' = y/t, y' = x, x(t_0) = a, y(t_0) = b.$$

Exercise 4.5

Find the values x_0 and a for which the Existence-Uniqueness Theorem guarantees existence and uniqueness for the IVP:

$$y' = y|x|, y(x_0) = a.$$

5. Exercises: Models and setting up ODEs

Exercise 5.1

What is the ODE of an object moving horizontally through a medium whose resistance is proportional to the object's velocity? Describe the long term behavior of the object.

Exercise 5.2

(a) In the theory of learning, the rate at which a subject is memorized is assumed proportional to the amount that is left to be memorized. Suppose M denotes the total amount of subject to be memorized and $A(t)$ is the amount memorized at time t . Set up a differential equation for $A(t)$. (b) Suppose in addition that the rate at which material is forgotten is proportional to $A(t)$. Set up a differential equation for $A(t)$.

Exercise 5.3

Suppose point T goes along the line $x = 1$ while dragging point P on the xy -plane by a string PT of length 1. Suppose T starts at $(1,0)$ and P at $(2,0)$. Find the path of P .

Exercise 5.4

Set up (don't solve) initial values problems for the following two situations: (a) An object is thrown up from a building of height h at 45 degrees with speed s ; (b) An object thrown travels for 2 seconds and then hits the ground at 45 degrees and speed s .

Exercise 5.5

Set up and solve the differential equation that describes the motion of an object of mass M placed on top of a spring with Hooke constant k standing vertically on the ground.

Exercise 5.6

(a) Describe the predator-prey model. (b) Set up a system of differential equations for the model and find its equilibria.

Exercise 5.7

Suppose an object is moving horizontally through a medium whose resistance is proportional to the

object's velocity:

$$v' = -ky, \quad k > 0.$$

Describe the long term behavior of the object. What if k grows with time?

Exercise 5.8

Derive the equations describing the predator-prey system and plot the phase portrait.

Exercise 5.9

Set up and solve the differential equation that describes the motion of an object of mass M suspended vertically by a spring with Hooke constant k .

Exercise 5.10

Set up (don't solve) initial values problems for the following two situations: (a) An object is thrown up from a building of height h at 45 degrees with speed s ; (b) An object thrown travels for 2 seconds and then hits the ground at 45 degrees and speed s .

Exercise 5.11

Set up and solve the differential equation that describes the motion of an object of mass M placed on top of a spring with Hooke constant k standing vertically on the ground.

Exercise 5.12

(a) Describe the predator-prey model. (b) Set up a system of differential equations for the model and find its equilibria.

Exercise 5.13

(a) In the theory of learning, the rate at which a subject is memorized is assumed proportional to the amount that is left to be memorized. Suppose M denotes the total amount of subject to be memorized and $A(t)$ is the amount memorized at time t . Set up a differential equation for $A(t)$. (b) Suppose in addition that the rate at which material is forgotten is proportional to $A(t)$. Set up a differential equation for $A(t)$.

6. Exercises: Qualitative analysis

Exercise 6.1

A sketch of the solution set of an ODE $y' = f(t, y)$ is shown below. What can you say about the sign of f ?

eventually.

Exercise 6.2

Suppose $y = 0$ is a stable (unstable) equilibrium of $y' = f(x, y)$. Suppose $g(x, y) = f(x, y)$ if $|y| < 1/x$. What can you tell about $y' = g(x, y)$?

Exercise 6.3

In the square $[-3, 3] \times [-3, 3]$, plot the direction field for the differential equation

$$y \frac{dy}{dx} = -x.$$

Sketch the stationary solution and three other solution curves.

Exercise 6.4

Sketch the direction field of the system:

$$x' = 2x - y, \quad y' = x - 3y$$

and several of its trajectories, identify the type of equilibrium.

Exercise 6.5

According to Newton’s law of cooling, the change of the temperature T of a body immersed in a medium of constant temperature A is described by the differential equation, with respect to time x :

$$\frac{dT}{dx} = k(A - T), \quad k > 0.$$

Solve it. Interpret your solution to explain why the temperature of the body will become equal to A ,

7. Exercises: Systems

Exercise 7.1

Sketch the trajectories of a system of linear ODEs $X' = FX$ if the matrix F has these pairs of eigenvalues and eigenvectors:

$$\lambda_1 = 2, V_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ and } \lambda_2 = -1, V_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Exercise 7.8

Suppose λ_1, λ_2 are the eigenvalues of a system of linear differential equations. Sketch the phase portrait of the system if (a) λ_1, λ_2 are complex with $Re\lambda_1 < 0$; (b) $\lambda_1 = \lambda_2$ with a single linearly independent eigenvector; (c) λ_1, λ_2 are real of opposite signs.

Exercise 7.2

Write the general solution of a system of linear ODEs $X' = FX$ if the matrix F has these pairs of eigenvalues and eigenvectors:

$$\lambda_1 = i, V_1 = \begin{bmatrix} 1 \\ i \end{bmatrix} \text{ and } \lambda_2 = -i, V_2 = \begin{bmatrix} 1 \\ -i \end{bmatrix},$$

and provide one non-trivial *real* solution of the system.

Exercise 7.3

Solve the system of differential equations:

$$x' = 2x - y, y' = x.$$

Exercise 7.4

Solve the system of linear equations:

$$x' = 4x - 5y, y' = 5x - 4y.$$

Exercise 7.5

Suppose λ_1, λ_2 are the eigenvalues of a system of linear ordinary differential equations. Sketch the phase portrait of the system if (a) λ_1, λ_2 are complex with $Re\lambda_1 < 0$; (b) $\lambda_1 = \lambda_2$ with a single linearly independent eigenvector; (c) λ_1, λ_2 are real of opposite signs.

Exercise 7.6

Solve the system

$$x' = -3x - 4y, y' = 2x + y.$$

Exercise 7.7

Solve the system of differential equations: $x' = 2x - y, y' = x$.

8. Exercises: Second order

Exercise 8.1

Carry out the substitution $y = Ce^{rt}$ to solve this ODE of second order: $y'' - 3y' + 2y = 0$.

Exercise 8.2

Find the general solutions to these differential equations with constant coefficients (1) $y'' - 6y' + 9y = 0$; (2) $2y'' - 5y' - 3y = 0$; (3) $y'' + 4y' + 7y = 0$.

Exercise 8.3

Find the general solutions to these differential equations with constant coefficients:

1. $y'' - 6y + 9y = 0$,

2. $2y'' - 5y' - 3y = 0$,

3. $y'' + 4y' + 7y = 0$.

Exercise 8.4

The differential equation $xy'' + y' = 0$ has a known solution $y_1 = \ln x$. Find another solution y_2 linearly independent from the first.

Exercise 8.5

The function $y_1 = e^x$ is a solution of the homogeneous equation

$$y'' - 3y' + 2y = 0.$$

Solve the non-homogeneous equation

$$y'' - 3y + 2y = 5e^{3x}.$$

Exercise 8.6

Given the differential equation $y'' - 2y + 2y = 0$, solve (a) the initial value problem

$$y(\pi) = 1, \quad y'(\pi) = 1,$$

and (b) the boundary value problem

$$y(0) = 1, \quad y(\pi) = 1.$$

Exercise 8.7

Solve the differential equation:

$$x^2y'' + (y')^2 = 0.$$

Exercise 8.8

Solve the initial value problem

$$(x - 1)y'' - xy + y = 0, \quad y(0) = -2, \quad y'(0) = 6,$$

in the form of a power series. (One extra point for representing the solution in a more compact form.)

Exercise 8.9

Solve the following differential equation:

$$y'' + 2y' + 4y = 0.$$

Exercise 8.10

Solve the differential equation

$$y'' + y'x = 0.$$

Exercise 8.11

Solve the following initial value problem:

$$y'' + 2y = 0, \quad y(0) = 0, \quad y'(0) = 1.$$

Exercise 8.12

Solve the following differential equation:

$$y'' + 2y' + 4y = 0.$$

9. Exercises: Advanced

Exercise 9.1

Linearize the following ODE at $x = 0$:

$$y' = \cos x - 1.$$

Do not solve.

Exercise 9.2

Solve the boundary value problem:

$$y'' - 4y = 0, \quad y(0) = 1, \quad y(1) = 2.$$

Exercise 9.3

Verify that the power series

$$y = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} x^n$$

is a solution of the differential equation

$$(x + 1)y'' + y' = 0.$$

Exercise 9.4

Given a differential equation

$$(\cos x)y'' + y = 0,$$

find two linearly independent solutions in the form of power series. Provide all terms up to x^2 . Recall that

$$\cos x = 1 - \frac{1}{2}x^2 + \frac{1}{4!}x^4 - \dots$$

Exercise 9.5

Provide the power series solution for

$$y' + y = 1.$$

Exercise 9.6

Suppose y_1, y_2 are two solutions of the equation

$$y'' + p(x)y' + q(x) = 0$$

and c_1, c_2 are constants. Show that $y = c_1y_1 + c_2y_2$ is also a solution of the equation.

Exercise 9.7

Find a general form of a particular solution of

$$y^{(3)} + y''' = 3e^x + 4x^2.$$

Exercise 9.8

By the power series method, solve the IVP

$$y'' + y' - 2y = 0, \quad y(0) = 1, \quad y'(0) = -2.$$

Exercise 9.9

Solve the following homogeneous equation:

$$2xy \frac{dy}{dx} = 4x^2 + 3y^2.$$

Exercise 9.10

Solve this exact equation

$$y^3 dx + 3y^2 x dy = 0.$$

Exercise 9.11

Solve the following differential equation in the complex domain:

$$y'' + y' + y = 0.$$

Exercise 9.12

Provide the power series solution for

$$y' + y = 1.$$

Exercise 9.13

Verify that the power series $y = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} x^n$ is a solution of the differential equation $(x = 1)y'' + y' = 0$.

Exercise 9.14

Given a differential equation $(\cos x)y'' + y = 0$, find two linearly independent solutions in the form of power series. Provide all terms up to x^2 . Recall that $\cos x = 1 - \frac{1}{2}x^2 + \frac{1}{4!}x^4 - \dots$

Exercise 9.15

Solve the initial value problem $(x - 1)y'' - xy' + y = 0, y(0) = -2, y'(0) = 6$, in the form of a power series. (One extra point for representing the solution in a more compact form.)

10. Exercises: PDEs

Exercise 10.1

Demonstrate that the function $u(t, x) = \sin t \cdot \sin x$ is a solution of the PDE:

$$u_{tt} = u_{xx} .$$

Suggest another solution.

Exercise 10.2

Compute the difference quotients with respect to x and y of the following function:

$y \mid x$	1	2	3	4	5	6
0	2	0	−1	1	2	−1
1	2	1	−2	1	1	−1
2	2	2	−3	1	0	−1
3	2	2	0	2	−1	−1
4	2	1	0	0	−2	−1
5	2	0	0	1	−3	−1

Exercise 10.3

Set up the difference equations for heat exchange between two adjacent objects and no exchange with the outside:

$x_{n+1} = \rule{1cm}{0.4pt}, y_{n+1} = \rule{1cm}{0.4pt}$

Exercise 10.4

Suppose a function $u(t, x, y)$ is used to represent the temperature at point (x, y) on the metal plate $[0, 1] \times [0, 1]$ at time t . (a) Express in terms of u the condition that the temperature of the edge of the plate is maintained at 0 at all times. (b) Express in terms of u the condition that at every location the temperature exponentially decays.

Exercise 10.5

Carry out $n = 4$ steps of the following difference equation with three cells, zero values outside, zero initial conditions, and your own choice of the time increment:

$x_{k+1,m} = x_{k,m} + (x_{k,m-1} + x_{k,m+1} + 1) \Delta t, m = 1, 2, 3 .$

11. Exercises: Computing

Computing assignment #1: Free fall. Carry out, or finish, or reproduce one of these projects from calculus 1 (the output should be a short presentation demonstrating an Excel spreadsheet accompanied by an explanation):

1. How do I throw a ball down a staircase so that it bounces off each step?
2. How should you throw a ball from the top of a 100 story building so that it hits the ground at 100 feet per second?
3. I would like to use a cannon with a muzzle velocity of 100 feet per second to bombard the inside of a fortification 300 feet away with walls 20 feet high.
4. I have a toy cannon and I want to shoot it from a table and hit a spot on the floor 10 feet away from the table.
5. How hard do I have to push a toy truck from the floor up a 30 degree incline to make it reach the top of the table at zero speed?
6. How fast does the shadow of a falling ball on a sliding ladder move?
7. How fast do I have to move my hand while spinning a sling in order to throw the rock 100 feet away?

Computing assignment #2: Euler’s method. For one of the ODEs below, subject it to Euler’s method analysis with Excel:

1. Analyze the domain of the right-hand side, apply the existence and uniqueness theorems.
2. Plot sufficiently many solutions.
3. Find patterns of the set of solutions, such as: periodicity, monotonicity, asymptotes, symmetry, and any other.

ODEs:

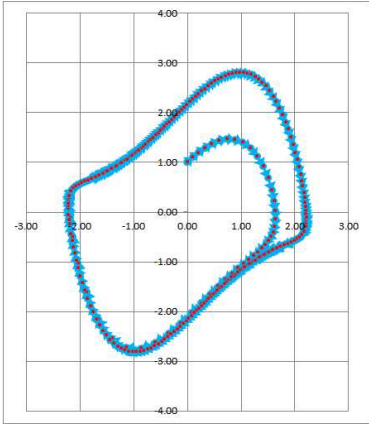
1. $y' = \tan(y^2)$
2. $y' = \cos(x + y)$
3. $y' = \cos y \cdot \sqrt{|y|}$
4. $y' = \sin y \cdot \sqrt{y}$
5. $y' = [x + y]$ (the **FLOOR** function in Excel)
6. $y' = \frac{1}{x - y}$

Computing assignment #3: Euler’s method on the plane. For one of the systems of ODEs below, perform the same tasks as in the last assignment:

1. $x' = \sin y, y' = \tan(y^2)$
2. $x' = |y|, y' = \cos(x + y)$
3. $x' = x + y, y' = \cos y \cdot \sqrt{|y|}$
4. $x' = y, y' = \sin y \cdot \sqrt{y}$
5. $x' = x - y, y' = [x + y]$ (the **FLOOR** function in Excel)
6. $x' = \cos(x + y), y' = \frac{1}{x - y}$

Computing assignment #4: Van der Pol oscillator. Apply the qualitative methods and Euler’s method to analyze the following second order equation:

$$x'' = m(1 - x^2)x' - x.$$



Index

- addition is preserved, [144](#)
- Algebra of Antiderivatives, [244](#)
- Algebra of Complex Numbers, [214](#)
- Algebra of Continuity, [233](#)
- Algebra of Derivatives, [244](#)
- Algebra of Limits of Functions, [232](#)
- Algebra of Limits of Sequences, [232](#)
- argument of complex number, [225](#)
- basis, [185](#)
- Basis on the Plane, [187](#)
- Bijections and Determinants, [167](#)
- cell, [379](#)
- cell decomposition, [29](#)
- cell function, [36](#)
- cells, [29](#)
- center, [303](#)
- Center of Mass Two-body System, [332](#)
- Chain Rule, [37](#)
- Characteristic Polynomial, [179](#)
- characteristic polynomial, [176](#)
- characteristic solution, [293](#)
- Classification of Linear Operators, [179](#)
- Classification of Linear Operators – Real Eigenvalues, [194](#)
- Classification of Linear Operators with Complex Eigenvalues, [240](#)
- Classification of Linear Systems I, [298](#)
- Classification of Linear Systems II, [303](#)
- Classification of Roots I, [217](#)
- Classification of Roots II, [218](#)
- Classification Theorem of Eigenvalues, [236](#)
- Columns are Values of Basis Vectors, [158](#)
- complex conjugate, [214](#)
- complex number, [213](#)
- Component-wise Convergence of Series, [245](#)
- Componentwise Convergence of Sequences, [222](#)
- composition, [209](#)
- Composition of Linear Operators, [209](#)
- Conservation of Energy, [320](#)
- Constant Approximation of ODEs, [97](#)
- Constant Multiple Rule for Complex Sequences, [223](#)
- Constant Multiple Rule for Series, [246](#)
- Continuity of IVP, [60](#)
- continuity of IVP, [55](#)
- Continuity of IVP for Location-Independent ODE, [56](#)
- Continuity of Polynomials, [233](#)
- continuous function, [233](#)
- convergent sequence, [222](#)
- converges absolutely, [245](#)
- Conversion of Complex Numbers, [226](#)
- derivative of a complex function, [242](#)
- Derivative of Polynomial, [244](#)
- determinant, [163](#), [235](#)
- Determinant Is Intrinsic, [168](#)
- Diff \Rightarrow Cont, [243](#)
- difference, [32](#)
- differential 0-form, [41](#)
- differential 1-form, [40](#)
- discrete form, [381](#)
- discrete forms, [30](#)
- discrete ODE of second order, [320](#)
- Discrete Second Kepler's Law, [328](#)
- dual cells, [120](#)
- dual forms, [121](#)
- eigenspace, [172](#), [235](#)
- Eigenspace Solutions, [292](#)
- eigenvalue, [170](#), [235](#)
- Eigenvalues and Eigenvectors, [176](#)
- Eigenvalues as Roots, [176](#)
- eigenvector, [170](#)
- eigenvectors, [171](#)
- Equilibrium of Linear System, [289](#)
- Error Bound, [86](#), [253](#)
- Euler solution, [76](#), [282](#), [289](#)
- Existence, [59](#), [279](#)
- existence, [52](#)
- Existence for Location-Independent ODE, [52](#)
- existence property, [277](#)
- face, [379](#)
- Fundamental Theorem of Calculus, [42](#)
- Fundamental Theorem of Discrete Calculus I, [34](#)
- Fundamental Theorem of Discrete Calculus II, [35](#)
- geometric series, [246](#)
- gluing, [377](#)
- heat equation, [383](#)
- identity operator, [149](#)
- Images of Lines, [150](#)
- imaginary numbers, [211](#)
- initial value problem, [52](#), [277](#), [289](#)
- Integer Power Formula, [243](#)
- integral of 1-form, [42](#)
- inverse matrix, [209](#)
- Inverse of Matrix of Dimension 2, [209](#)
- IVP, [277](#)
- limit of a function, [232](#)
- Line Collapses, [167](#)
- Linear Approximation of ODEs, [100](#)
- Linear ODE, [63](#)

- linear ODEs of second order, 114
- linear operator, 147–149, 158–160
- Linear Operator at 0, 148
- Linear Operator in Terms of Basis, 190
- Linear Operators and Linear Combinations, 147
- Linear Operators vs. Matrices, 147
- Locality, 232
- Locations in Two-body System, 332
- Matrix of Composition, 209
- Matrix of Rotation, 160
- maximal solution, 60
- Mixed Second Derivative of Wave, 402
- modulus of complex number, 225
- Monotonicity of Solutions, 89
- Multiples of Eigenvectors, 171
- Multiplication of Complex Numbers, 227
- Newton’s Law of Cooling, 353, 365
- Non-homogeneous Linear ODE, 67
- Non-positive Discriminant, 240, 241
- Non-zero Determinant, 235
- Non-Zero Solutions, 165, 167
- ODE, 47
- ODEs of second order, 114
- One-sided Error Bound, 88
- One-to-one Linear Operator, 162
- parametric curve of the uniform motion, 257
- parametric curve of uniformly accelerated motion, 261
- Planetary Motion in Polar Coordinates, 329
- power series, 248
- Preimages of Zero, 180
- Preserving Addition, 144
- Preserving Scalar Multiplication, 146
- primal and dual domains, 119
- product of matrix and vector, 130
- Product Rule for Complex Sequences, 224
- Quotient Rule for Complex Sequences, 225
- Radius of Convergence, 249
- radius of convergence, 249
- Representation, 191
- Representation In Terms of Eigensolutions, 293
- saddle, 298
- sampling, 42
- scalar multiplication is preserved, 146
- second difference, 121
- second difference quotient, 122
- sequence of partial sums, 245
- sequence tends to infinity, 222
- Shifted Solutions, 276
- singular matrix, 163
- Singular Matrix and Determinant, 164
- solution of ODE, 48
- solution of system of ODEs, 276, 289
- Solution of Wave Equation, 403
- Solutions of Heat PDE, 373
- Solutions of Heat PDE Dimension 2, 393
- Solutions of Heat PDE With Zero Boundary Condition, 374
- Solutions of Separable ODEs, 62
- spreadsheet, 396
- stable and unstable focus, 302
- stable node, 297
- standard form of complex number, 214
- Stationary Solutions, 89
- sum, 33
- Sum of Geometric Series, 247
- sum of series, 245
- Sum of Solutions of the Heat PDE, 374
- Sum Rule for Complex Sequences, 223
- Sum Rule for Series, 246
- Term-by-Term Differentiation and Integration, 250
- Time Independent ODEs, 88
- Trace and Discriminant, 197
- trace of matrix, 178
- transformation of the plane, 133
- Uniqueness, 60, 279
- uniqueness, 53
- Uniqueness for Location-Independent ODE, 54
- Uniqueness of Limit, 222
- Uniqueness of Power Series, 250
- Uniqueness of Sum, 245
- uniqueness property, 277
- Values of Basis Vectors Are Columns, 159
- vector-valued discrete form, 306
- weak solution of ODE, 48
- Weierstrass M-Test, 248
- zero operator, 148